DOCUMENT RESUME

ED 093 452 PS 007 109

AUTHOR	Cain, Glen G.; Barnow, Burt S.
TITLE	The Educational Performance of Children in Head Start and Control Groups. Final Report.
INSTITUTION	Wisconsin Univ., Madison.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.; Office of Economic Opportunity, Washington, D.C.
BUREAU NO	BR-2-0721
PUB DATE	24 Sep 73
GRANT	OEG-0-72-1384
NOTE	273p.; The text of this report was originally submitted by Burt S. Barnow, as a Ph.D. Dissertation, University of Wisconsin, 1973
EDRS PRICE	MF-$0.75 HC-$12.60 PLUS POSTAGE
DESCRIPTORS	*Evaluation Methods; *Evaluation Techniques; *Research Methodology; *Research Problems; *Statistical Analysis
IDENTIFIERS	*Project Head Start

ABSTRACT

This report is a re-analysis of the data collected and analyzed by the Westinghouse Learning Corporation (WLC) and Ohio University concerning the average impact of Project Head Start on the cognitive development of a nationwide sample of children. The re-analysis was considered necessary because of unclear methodological issues in evaluation research and the existence of much data that was not used in the WLC analysis. A basic question raised is whether the WLC data can be used to produce unbiased estimates of the effects of Head Start. Several formal models of Head Start Evaluation are presented in order to determine the conditions that would lead to biased and unbiased estimates. The WLC report did not describe its procedures for selecting children for Head Start and control groups. For the re-analysis, the following modifications of the WLC study were made: (1) use of ungrouped instead of grouped data, (2) expanded list of socioeconomic and demographic indendent variables, and (3) inclusion of the Head Start variable in a manner to allow for different effects for children from various ethnic groups and family structures. Findings generally support the original study, but indicate that Head Start was more effective than the WLC report indicated. (DP/Author)

ED 093452

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: UD | PS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. ~~UD~~
transfer

FINAL REPORT

PROJECT NO. ~~5-0198~~ 2-0721
GRANT NO. OEG-0-72-1384

THE EDUCATIONAL PERFORMANCE OF CHILDREN IN HEAD START AND CONTROL GROUPS

GLEN G. CAIN
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN
53706

SEPTEMBER 24, 1973

The research reported herein was performed pursuant to a grant with the National Institute of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U. S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

OFFICE OF RESEARCH GRANTS

PS 007109

Summary of Report on The Educational Performance
of Children in Head Start and Control Groups
(Project No. ~~50196~~, Grant No. OEG-0-72-1384
2-0721
Principal Investigator, Glen G. Cain

Co-Investigator, Burt S. Barnow

PS 007109

I. Background

This report is a re-analysis of the data collected and analyzed by the Westinghouse Learning Corporation (WLC) and the Department of Educational Psychology at Ohio University. The WLC study was the first evaluation of a nationwide sample of Head Start centers, and it measured the average impact of the program on the cognitive development of children. The study generally concluded that Head Start had no significant effect on the test scores of the participating children. This conclusion and the methodology were sharply criticized as soon as the findings were made public, and the controversy has continued since then.

A re-examination of the study and data is very much in order. There are several methodological issues in evaluation research which were not fully clarified by the WLC study and the ensuing debates. Also, much of the data collected, particularly that pertaining to parental and Head Start Center characteristics, were not used. As a consequence, opportunities were missed both to refine the estimate of the effect of Head Start and to estimate the effects of a number of variables which are interesting in their own rights. Our re-analysis has attempted to rectify these gaps and shortcomings.

The empirical results we report generally support those of the original study, although there are additional results which qualify the pessimistic findings about compensatory education programs. At the same time our analysis of the methods of the WLC study and of the theory underlying the estimation model make clear how the data and study design severely limit the confidence with which the empirical results can be accepted. In this respect, the criticisms of the study, in modified form, are supported.

II. Principal Findings

1. Methodological issues in evaluation research.

The methodological problem of obtaining unbiased measures of a treatment effect in the absence of a controlled experiment is sharply posed in the WLC study. The problem is made more difficult because the study was forced to deal with data collected after the Head Start programs had been completed, so no pre-program test scores of educational achievement were available. Instead, the design was based on the following procedure of data collection. Neighborhoods where Head Start Centers operated were randomly sampled and children who did and did not participate in the programs -- matched by age, race, sex, and prior kindergarten attendance -- were randomly selected and tested. This sampling process was replicated for children in the first, second, and third grades and for a smaller sample of children who attended full-year Head Start programs.

The crucial question is the comparability of the Head Start and control groups. The WLC investigators implicitly argue that the groups are comparable in their ability (or, more narrowly, test-taking ability), at least after controlling for a measure of socio-economic status of the parents. The criticisms of the study, particularly the well-known article by D.T. Campbell and A. Erlebacher ("How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look

Harmful"), assert that the control group had a higher mean ability and that, as a consequence, there is a downward bias in the treatment effect.

The critics raise the methodological point that a difference in mean ability between the treatment and control population will produce a bias in the treatment effect, despite matching or regression methods. We show that a bias is not a necessary consequence of the differences in population means. In principle, a bias will not result when the basis for allocating the subjects to the two groups is known and the allocation information is used in the regression model. This method of modeling the selection procedure is an alternative to a true experiment (with randomization).

We argue that this alternative strategy is not (or should not) be a special case with no practical significance. The selection procedures are known (or knowable) to the program administrators, since they have control over the procedures. The procedures may consist of using pre-test scores, per capita income of the family, or some combination of these and other characteristics as criteria for selections and assignments. Such procedures will probably result in the treatment and control groups having different ability means. However, this disparity in the groups will not produce a bias in the treatment effect as long as the selection variables are included in the model. It is clear, however that the selection procedures used by the WLC study are unknown, and therefore the bias issue cannot be resolved for these data.

2. Empirical Findings

Regression analysis is used in the reanalysis of the data. The statistical model is similar to the one used in the Westinghouse study, but several modifications have been made. The major changes are that we use ungrouped rather than grouped data to increase efficiency, and that we expand the list of socioeconomic and demographic independent variables to cover all available variables which should be included in the educational production function to reduce the possible bias and to learn about the educational production process. Also, the Head Start variable is included in a manner to allow for different effects for children from various ethnic groups and family structures.

The findings of the reanalysis are compatible with those of the Westinghouse study but differ because of the changes in the structure of the model. Both summer and full-year Head Start appear to be effective for white children from mother-headed families but ineffective for whites from two-parent families; a gain equivalent to at least 5 IQ points was found for the white children from mother-headed families who were in the first, second, and third grades at the time of the study. For all balck children there was a Head Start effect of about 5 IQ points for the first grade samples, but there was a zero effect for the second and third grade samples. The zero effects may not be indicative of a fading of the immediate impact because the children in the three grades were not the same ones and the Head Start programs may have changed over the three-year period. For both races there was no significant difference found between the effects of full-year and summer programs. Discriminant analysis was employed to determine if the Head Start and control groups differ on the socioeconomic variables available; for most samples there was an insignificant difference in favor of the controls.

The implications of the findings are that Head Start may be effective for specific types of disadvantaged children--whites from mother-headed families and children from minority ethnic groups. However, we have not found full-year Head Start programs to be significantly more effective than summer programs; thus the Westinghouse recommendation that summer programs be replaced by full-year programs appears unwarranted.

3. Further Results

As mentioned above, use was made of discriminant analysis to measure the differences between the Head Start and control children. A refinement of the technique for discriminating between the two groups was to obtain the predicted test score for Head Start children, given *their characteristics* -- age, sex, socio-economic background, etc. -- on the assumption that these characteristics have the same effects on the test scores as were estimated for the control group. This method provides one measure of how similar or different the two groups are in the relevant metric of the analysis; i.e. test scoring ability. The Head Start children were found to be lower in predicted mean ability, as measured by test scores, by about five percent -- in one test by an amount equivalent to 3.8 I.Q. points.

Further analysis is given of the effects on educational achievement and on the Head Start effect on educational achievement of a variety of socio-economic and demographic variables: race, Mexican-American ethnicity, socio-economic status of the parents, one parent families, and mothers working. Except for race and single-parentness, few variables had any significant interaction effect with Head Start. All the variables usually had an additive effect in the expected direction on test score performance, and it is interesting to note that the educational and occupational characteristics of the mother had the largest and most significant effects among socio-economic variables. Finally, we report results of the effects of kindergarten experience and assess the generally favorable effect on test scores of this variable.

ED 093452

THE EFFECTS OF HEAD START AND SOCIOECONOMIC STATUS ON COGNITIVE DEVELOPMENT OF DISADVANTAGED CHILDREN

BY

BURT S. BARNOW

A thesis submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY
(Economics)

at the

UNIVERSITY OF WISCONSIN

1973

PS 007109

© Copyright by Burt S. Barnow 1973

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Burt S. Barnow

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

To my wife Renee

TABLE OF CONTENTS

ABSTRACT vi

ACKNOWLEDGMENTS ix

LIST OF TABLES xi

1. INTRODUCTION 1

2. STATISTICAL MODELS FOR HEAD START EVALUATION 7
   - Introduction 7
   - A Model with No Errors in the Independent Variables 12
   - The Campbell-Erlebacher Two-Population Model 15
   - A One-Population Model with Nonrandom Selection 19
   - A One-Population Model with Interaction 24
   - Selection on the Basis of Pretest Scores 27
   - An Omitted Variable Model 35
   - Application of the Models to the Westinghouse Study 39

3. REVIEW OF THE WESTINGHOUSE STUDY 45
   - Introduction and Background 45
   - Plan of Research 47
   - Instruments Used in the Westinghouse Study 51
   - Statistical Model 58
   - Results 61
   - Conclusions of the Westinghouse Study 66

4. CRITICISMS OF THE WESTINGHOUSE STUDY 70
   - Introduction 70
   - The Focus of the Westinghouse Study 72
   - The Instruments Used in the Westinghouse Study 75
   - Sampling Procedures 78
   - Criticisms of the Statistical Analysis 82
   - Interpretations of Statistical Findings and Strategies for Evaluations 90
   - Summary 95

5. EMPIRICAL FINDINGS 98
   - Introduction 98
   - Hypotheses 104
   - Empirical Findings for the Primary Samples 110
   - Summary of Head Start and Kindergarten Effects for All Samples 124
   - Techniques for Measuring Head Start and Control Pretreatment Differences 126
   - Supplementary Models 136
   - Summary of the Appendix 157
   - Comparison of the Reanalysis to the Westinghouse Findings 158

6. SUMMARY AND CONCLUSIONS 163
- Summary of Findings 163
- Comparison with Other Evaluations of Preschool Programs 171
- Policy Implications 173

APPENDIX: ADDITIONAL EMPIRICAL RESULTS 178
- Analyses of the Subtests for the First Grade, Summer, Both Parents Sample 178
- Regressions of Mean Test Scores on Individual Characteristics for All Samples 180
- Head Start Discriminant Analyses 183
- Kindergarten Discriminant Analyses 184

BIBLIOGRAPHY 247

Abstract

THE EFFECTS OF HEAD START AND SOCIOECONOMIC STATUS ON COGNITIVE DEVELOPMENT OF DISADVANTAGED CHILDREN

Burt S. Barnow

Under the supervision of Professor Glen G. Cain

The Westinghouse Learning Corporation-Ohio University study was the first evaluation of Head Start carried out on a national scale to measure the average impact of Head Start upon the cognitive and affective development of children. The study found that summer programs are ineffective and full-year programs are only marginally effective in raising the cognitive and affective development of preschool children. The evaluation has received much criticism. This thesis reviews the controversy about the study and reanalyzes the data in an economic framework by attempting to incorporate Head Start into an educational production function.

One of the major areas of controversy for the study concerns whether the Westinghouse data can be used to produce unbiased estimates of the effects of Head Start. The thesis presents several formal models of Head Start evaluation to determine conditions which will lead to biased and unbiased estimates. It is determined that specification errors in the model (such as fallible measures of independent variables, or unobservable variables) will lead to biased estimates under certain procedures for selecting children for the Head Start and control groups. As the selection procedures used by the Westinghouse study are unknown, the bias issue cannot be resolved for the Westinghouse data.

Regression analysis is used in the reanalysis of the data. The statistical model is similar to the one used in the Westinghouse study, but several modifications have been made. The major changes are that we use ungrouped rather than grouped data to increase efficiency, and that we expand the list of socioeconomic and demographic independent variables to cover all available variables which should be included in the educational production function to reduce the possible bias and to learn about the educational production process. Also, the Head Start variable is included in a manner to allow for different effects for children from various ethnic groups and family structures.

The findings of the reanalysis are compatible with those of the Westinghouse study but differ because of the changes in the structure of the model. Both summer and full-year Head Start appear to be effective for white children from mother-headed families but ineffective for whites from two-parent families; a gain equivalent of at least 5 IQ points was found for the white children from mother-headed families who were in the first, second, and third grades at the time of the study. For all black children there was a Head Start effect of about 5 IQ points for the first grade samples, but there was a zero effect for the second and third grade samples. The zero effects may not be indicative of a fading of the immediate impact because the children in the three grades were not the same ones and the Head Start programs may have changed over the three-year period. For both races there was no significant difference found between the effects of full-year and summer programs. Discriminant analysis was employed to determine if the Head Start and control groups differ on the socioeconomic variables available; for most samples there was no significant difference.

The implications of the findings are that Head Start may be effective for specific types of disadvantaged children--whites from mother-headed families and children from minority ethnic groups. However, we have not found full-year Head Start programs to be significantly more effective than summer programs; thus the Westinghouse recommendation that summer programs be replaced by full-year programs appears unwarranted.

ACKNOWLEDGMENTS

Financial support for this study was provided in part by the National Institute of Education, U.S. Department of Health, Education, and Welfare pursuant to Grant No. OEG-0-72-1384, and by the Institute for Research on Poverty, University of Wisconsin-Madison, through funds granted by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. The author wishes to express his gratitude to these organizations and to note that the opinions expressed in this study do not necessarily represent those of the sponsoring agencies. Nor do the opinions represent those of the following people whose guidance I gratefully acknowledge. I take sole responsibility for all the material presented.

I am especially grateful to Professor Glen G. Cain who suggested this topic of research and has provided a great deal of guidance and many helpful suggestions throughout the study. I have also benefited from the suggestions of my other committee members, Professors Arthur S. Goldberger and Robert H. Haveman. I would also like to thank Professor Victor G. Cicirelli for providing the data as well as for providing suggestions on my research. Professors Donald T. Campbell, Robert A. Hauser, and Burt Fisher have also been helpful. To all who have provided assistance I wish to express my appreciation. Finally, I would like to express my appreciation to Cathy Ersland and Jean Arnold for their aid in typing the manuscript. My greatest gratitude is to

Renee Barnow who has not only typed much of this manuscript, but whose editing and encouragement have proved to be an invaluable help.

LIST OF TABLES

2.1 Description of Variables Used in the Theoretical Models 11

5.1 Description of Variables Used in the Reanalysis 100

5.2 Means and Standard Deviations for Grade 1, Summer, Both Parents Present Sample 111

5.3 Means and Standard Deviations for Grade 1, Full-Year, Both Parents Present Sample 113

5.4 Effects of Individual Characteristics for Grade 1, Both Parents Present Sample, on Child's ITPA Score, Summer and Full Year 115

5.5 Effects of Individual Characteristics for Grade 1, Both Parents Present Sample, on Child's MRT Score, Summer and Full Year 117

5.6 Summary of Effects of Head Start and Kindergarten for All Regressions with ITPAMN as the Dependent Variable 125

5.7 Discriminant Regression Results for Grade 1, Both Parents Samples, with Head Start as the Dependent Variable, Summer and Full Year 129

5.8 Discriminant Regression Results for Grade 1, Both Parents Samples, with Kindergarten as the Dependent Variable, Summer and Full Year 132

5.9 Description of Head Start Center Characteristic Variables 140

5.10 Effects of Individual and Center Characteristics for Grade 1, Both Parents Present, on Child's ITPA Score, Summer and Full Year 143

5.11 Effects of Individual and Center Characteristics for Grade 1, Both Parents Present, on Child's MRT Score, Summer and Full Year 147

5.12 Summary of Head Start Coefficients when Data is Stratified by Employment Status of the Mother 155

A-1 Correlation Matrix of All Test Scores for Grade 1, Summer, Both Parents Sample 185

A-2 Effects of Individual Characteristics on ITPA Mean 189

A-3 Effects of Individual Characteristics on ITPA Auditory Reception Subtest 190

A-4 Effects of Individual Characteristics on ITPA Visual Reception Subtest 191

A-5 Effects of Individual Characteristics on ITPA Visual Sequential Memory Subtest 192

A-6 Effects of Individual Characteristics on ITPA Auditory Association Subtest 193

A-7 Effects of Individual Characteristics on ITPA Auditory Sequential Memory Subtest 194

A-8 Effects of Individual Characteristics on ITPA Visual Association Subtest 195

A-9 Effects of Individual Characteristics on ITPA Visual Closure Subtest 196

A-10 Effects of Individual Characteristics on ITPA Verbal Expression Subtest 197

A-11 Effects of Individual Characteristics on ITPA Grammatic Closure Subtest 198

A-12 Effects of Individual Characteristics on ITPA Manual Expression Subtest 199

A-13 Effects of Individual Characteristics on MRT Mean 200

A-14 Effects of Individual Characteristics on MRT Word Meaning Subtest 201

A-15 Effects of Individual Characteristics on MRT Listening Subtest 202

A-16 Effects of Individual Characteristics on MRT Matching Subtest 203

A-17 Effects of Individual Characteristics on MRT Alphabet Subtest 204

A-18 Effects of Individual Characteristics on MRT Numbers Subtest 205

A-19 Effects of Individual Characteristics on MRT Copying Subtest 206

A-20 Effects of Individual Characteristics for Grade 1, Full-Year, Both Parents Sample on Child's ITPA Score 207

A-21 Effects of Individual Characteristics for Grade 1, Full-Year, Both Parents Sample on Child's MRT Score 208

A-22 Effects of Individual Characteristics for Grade 1, Full-Year, Mother Only Sample on Child's ITPA Score 209

A-23 Effects of Individual Characteristics for Grade 1, Full-Year, Mother Only Sample on Child's MRT Score 210

A-24 Effects of Individual Characteristics for Grade 1, Summer, Both Parents Sample on Child's ITPA Score 211

A-25 Effects of Individual Characteristics for Grade 1, Summer, Both Parents Sample on Child's MRT Score 212

A-26 Effects of Individual Characteristics for Grade 1, Summer, Mother Only Sample on Child's ITPA Score 213

A-27 Effects of Individual Characteristics for Grade 1, Summer, Mother Only Sample on Child's MRT Score 214

A-28 Effects of Individual Characteristics for Grade 2, Full-Year, Both Parents Sample on Child's ITPA Score 215

A-29 Effects of Individual Characteristics for Grade 2, Full-Year, Both Parents Sample on Child's SAT Score 216

A-30 Effects of Individual Characteristics for Grade 2, Full-Year, Mother Only Sample on Child's ITPA Score 217

A-31 Effects of Individual Characteristics for Grade 2, Full-Year, Mother Only Sample on Child's SAT Score 218

A-32 Effects of Individual Characteristics for Grade 2, Summer, Both Parents Sample on Child's ITPA Score 219

A-33 Effects of Individual Characteristics for Grade 2, Summer, Both Parents Sample on Child's SAT Score 220

A-34 Effects of Individual Characteristics for Grade 2, Summer, Mother Only Sample on Child's ITPA Score 221

A-35 Effects of Individual Characteristics for Grade 2, Summer, Mother Only Sample on Child's SAT Score 222

A-36 Effects of Individual Characteristics for Grade 3, Summer, Both Parents Sample on Child's ITPA Score 223

A-37 Effects of Individual Characteristics for Grade 3, Summer, Both Parents Sample on Child's SAT Score 224

A-38 Effects of Individual Characteristics for Grade 3, Summer, Mother Only Sample on Child's ITPA Score 225

A-39 Effects of Individual Characteristics for Grade 3, Summer, Mother Only Sample on Child's SAT Score 226

A-40 Discriminant Regression Results for Grade 1, Full-Year, Both Parents Sample, with Head Start as the Dependent Variable 227

A-41 Discriminant Regression Results for Grade 1, Full-Year, Mother Only Sample, with Head Start as the Dependent Variable 228

A-42 Discriminant Regression Results for Grade 1, Summer, Both Parents Sample, with Head Start as the Dependent Variable 229

A-43 Discriminant Regression Results for Grade 1, Summer, Mother Only Sample, with Head Start as the Dependent Variable 230

A-44 Discriminant Regression Results for Grade 2, Full-Year, Both Parents Sample, with Head Start as the Dependent Variable 231

A-45 Discriminant Regression Results for Grade 2, Full-Year, Mother Only Sample, with Head Start as the Dependent Variable 232

A-46 Discriminant Regression Results for Grade 2, Summer, Both Parents Sample, with Head Start as the Dependent Variable 233

A-47 Discriminant Regression Results for Grade 2, Summer, Mother Only Sample, with Head Start as the Dependent Variable 234

A-48 Discriminant Regression Results for Grade 3, Summer, Both Parents Sample, with Head Start as the Dependent Variable 235

A-49 Discriminant Regression Results for Grade 3, Summer, Mother Only Sample, with Head Start as the Dependent Variable 236

A-50 Discriminant Regression Results for Grade 1, Full-Year, Both Parents Sample, with Kindergarten as the Dependent Variable 237

A-51 Discriminant Regression Results for Grade 1, Full-Year, Mother Only Sample, with Kindergarten as the Dependent Variable 238

A-52 Discriminant Regression Results for Grade 1, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable 239

A-53 Discriminant Regression Results for Grade 1, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable 240

A-54 Discriminant Regression Results for Grade 2, Full-Year, Both Parents Sample, with Kindergarten as the Dependent Variable 241

A-55 Discriminant Regression Results for Grade 2, Full-Year, Mother Only Sample, with Kindergarten as the Dependent Variable 242

A-56 Discriminant Regression Results for Grade 2, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable 243

A-57 Discriminant Regression Results for Grade 2, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable 244

A-58 Discriminant Regression Results for Grade 3, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable 245

A-59 Discriminant Regression Results for Grade 3, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable 246

Chapter 1

Introduction

> So crucial is the matter of early growth that we must make a national commitment to providing all American children an opportunity for healthful and stimulating development during the first 5 years of life. (Richard Nixon, February 19, 1969)

With these words President Nixon expressed his support for the concept of early intervention for children from disadvantaged backgrounds. One of the principal programs aimed at aiding these children has been Head Start. Head Start is a national preschool education program whose purpose has been to prepare children from disadvantaged backgrounds for entrance into formal education in the primary grades. The philosophy underlying the program is that one reason children from disadvantaged backgrounds perform poorly in school, and hence drop out and remain impoverished, is that the home environment does not provide the stimulation and amenities found in middle-class homes. By intervening between the ages of three and five, the program seeks to give these children a "head start" in their attitudes and cognitive development, and thus break out of the cycle of poverty. Originally proposed as a pilot project in President Johnson's "War on Poverty" in 1965, the program was greeted with such astounding popularity that the funding for the first summer programs was increased from $17 million to $103 million according to Sar A. Levitan (1969, p. 136). The program has continued to grow, both in popularity and in size, and it remains as one of the few remnants of the Johnson Administration's antipoverty program to retain widespread support.

Although Head Start has been sold to the public mainly on its contributions to the cognitive development of children, the program has stressed all aspects of children's growth in order to help its participants break out of the cycle of poverty. Edith H. Grotberg (1969, p. 1) reports that the seven objectives of Head Start are:

A. Improving the child's physical health and physical abilities

B. Helping the emotional and social development of the child by encouraging self-confidence, spontaneity, curiosity, and self-discipline

C. Improving the child's mental processes and skills with particular attention to conceptual and verbal skills

D. Establishing patterns and expectations of success for the child which will create a climate of confidence for his future learning efforts

E. Increasing the child's capacity to relate positively to family members and others while at the same time strengthening the family's ability to relate positively to the child and his problems

F. Developing in the child and his family a responsible attitude toward society, and fostering constructive opportunities for society to work together with the poor in solving their problems

G. Increasing the sense of dignity and self-worth within the child and his family

To meet these objectives certain patterns have been established for Head Start programs. Most centers provide physical and dental check-ups to improve the physical health of the children, and, in addition, many centers have provided treatment for physical maladies. Most programs also provide one or two balanced meals for the children who attend. Parent participation has been considered an important part of the program, and many centers encourage parents to serve on the paid staff or as volunteers. The focus on the intellectual development of children is what has differentiated Head Start from a custodial day care program, and has been one of the primary reasons for its continued popularity.

In 1968, the Office of Economic Opportunity awarded a contract to the Westinghouse Learning Corporation and Ohio University to assess the average impact of Head Start on the cognitive and affective development of children who had participated in the program during its first three years. The Westinghouse report was released a year later and created a major controversy by concluding that summer Head Start was ineffective, and full-year programs were only marginally effective. Although the Westinghouse study was not intended to be a definitive evaluation of Head Start and several more refined evaluations were commissioned at the same time, the Westinghouse study is the only major national evaluation that has been completed at this time. The negative findings of the study have cast doubt on the usefulness of Head Start as a tool by which disadvantaged children can increase their cognitive development--and eventually increase their earning potential and break out of the cycle of poverty.

The Westinghouse study has been criticized by government officials, Head Start officials, and academicians. Various critics have claimed that the study asked the wrong questions, the sampling procedures used were incorrect, the statistical analysis was incorrect, and the interpretations of the analysis were unsound. The goal of this dissertation is to reexamine the controversy over the Westinghouse report and to reanlayze the data in order to make better estimates of the effects of Head Start on the cognitive development of children. Although enhancing cognitive development is only one of the many objectives of Head Start, we have concentrated on this particular aspect because it is the primary one that distinguishes Head Start from day care programs. We do not present a cost-benefit analysis because we are not qualified to place a dollar value on cognitive benefits to preschool children.

Although the evaluation of educational programs has traditionally been in the realm of psychology, in recent years economists have become interested in the field. Education can be viewed as an investment in human beings, and a great deal of research has been done by economists in the last ten years in the area of what is called human capital. Preschool education programs are similar in many ways to the manpower training programs that have traditionally been studied by economists. In both types of programs the participant receives a type of training which enhances his stock of human capital; in the case of manpower training programs the participant learns new skills which presumably increase his ability to earn income, and a similar argument can be made for education. For training programs and preschool education, scarce resources are used in the process. The items of interest to an economist are finding the most efficient means of producing the training and learning whether or not the benefits of the training exceed the costs to make the program worthwhile.

Education can thus be viewed as an economic good that provides a stream of benefits to the recipient. When viewed in this manner it is logical to try to determine what the input factors are in the production of education. Samuel Bowles (1970, p. 12) defines an educational production function as "the relationship between school and student inputs and a measure of school output." Thus, one of the goals of this dissertation is to determine how Head Start should be included in the educational production function for primary grade children. Research concerning educational production functions is still in the early stages, and there are problems in determining which variables should be included and what functional forms should be used. We have made simplifying assumptions where necessary in order to carry out our reanalysis of the Westinghouse data. As Bowles (1970, p. 19) notes, "The dearth of knowledge concerning the learning process makes any a priori specification of form for the educational production relationships particularly difficult." We have followed Bowles in using a linear functional form in our analysis, but we realize that substantial research remains to be done in determining the appropriate functional form.

The dissertation has been divided into six chapters and an appendix. In Chapter 2 the statistical problems involved in an evaluation of Head Start are discussed in the context of several formal models; we demonstrate how measurement problems and the selection procedure used to assign children to the Head Start and control groups can sometimes lead to biased estimates of the treatment effect. Chapter 3 reviews the history of the Westinghouse study and includes a description of the instruments used in the analysis, the methods of statistical analysis, the major findings of the study, and the interpretations and policy recommendations of the Westinghouse researchers. We review the criticisms found in the literature about the Westinghouse study and present original criticisms in Chapter 4; in this chapter we also

PS 007109

outline a procedure to be used for making policy inferences from the evaluation. Our reanalysis of the Westinghouse data and interpretations of the findings are in Chapter 5. Chapter 6 summarizes our findings, compares our findings to those of other evaluations of preschool programs, and offers policy recommendations. The Appendix includes supplementary empirical analyses that were not considered as important as those included in Chapter 5; a summary of the Appendix is included in Chapter 5.

Chapter 2

Statistical Models for Head Start Evaluation

1. Introduction

Evaluations of social action programs such as Head Start are often complicated by the problems inherent in such experiments. Some of the most common problems encountered include: ignorance of the structural form of the appropriate model, errors in the measurement of one or more of the independent variables, and unobservable variables that should be included in the model. In this chapter several models for evaluating programs such as Head Start will be considered, and the consequences of various specification errors in the empirically testable analogues of these models are determined. We then determine under what conditions regression analysis will lead to unbiased estimates of the treatment effect of participating in Head Start, and what direction the bias will take when it is impossible to get unbiased estimates.

Because we consider models which determine the effects of Head Start upon the cognitive development of children, we must define "cognitive development." We shall use the term in a broad sense to mean the mastery of certain skills, processes, concepts, and facts. Psychologists often distinguish between the concepts of intelligence and achievement, where intelligence refers to one's capacity to master skills, concepts, processes, and facts, and where achievement refers to the level of mastery of these items.

The matter of determining the appropriate measure of cognitive development is complicated further if we consider the possibility that intelligence and achievement can be multidimensional phenomena. There is no reason why achievement cannot be divided into many specific areas which can then be further subdivided; i.e., the broad area of mathematics can be divided into subdivisions such as algebra, geometry, and calculus, and each of these can be divided into even more specific areas. Although intelligence is often treated as a unidimensional characteristic, some psychologists and geneticists have tried to decompose it. Jensen (1968, p.56), for example approvingly cites the work of Fifer (1965) who decomposed general intelligence into verbal, reasoning, number, and spatial abilities. Jencks et al. (1972, p.55) argue that because the five tests used in the Equality of Educational Opportunity Survey (EEOS), also known as the Coleman Report, are highly correlated with each other, scholastic ability is a one-dimensional concept. But, the tests used in the EEOS study may have been achievement-oriented and the high correlations may have been a result of high correlations in the instruction received in the different areas.

The problems included in the definition and measurement of cognitive development will not be considered in this thesis--this area is more appropriate for study by educational psychologists. In this chapter it will be assumed that cognitive development is a unidimensional variable or that we are only interested in one dimension of it. In the empirical section of the thesis, regression analysis will be undertaken using all of the measures available from the Westinghouse study.

In the remainder of this chapter we shall examine the possibilities of carrying out an unbiased evaluation of Head Start by using regression analysis. In particular, we shall examine quasi-experimental situations where random assignment was not used and/or where an ex post facto analysis must be used because there is no pretreatment information available. This does not imply that quasi-experimental analyses are more desirable than true experiments, but rather that when a quasi-experimental analysis is the only feasible means of carrying out an evaluation, the analysis may not lead to bias in the estimates of treatment effect. Thus, we shall demonstrate that the following statement by Campbell and Erlebacher (1970, p.185) is misleading:

> Evaluations of compensatory educational efforts such as Head Start are commonly quasi-experimental or *ex post facto*. The compensatory program is made available to the most needy, and the "control" groups then sought from among the untreated children in the same community. Often this untreated population is on the average more able than the "experimental" group. In this situation, the usual procedures of selection, adjustment, and analysis produce systematic biases in the direction of making the compensatory program look deleterious.

To prove our points, several models with various relationships among the appropriate variables for an evaluation of Head Start will be presented. The models will then be examined to determine whether or not regression analysis will give unbiased estimates of treatment effects for the population. (It should be noted that regression analysis is equivalent to analysis of covariance so that our results could be expressed equally well in terms of analysis of covariance.) The assumptions that are made in each of the models are often crucial in determining if an unbiased evaluation can be carried out. Relaxation of some of these assumptions can lead to important changes in the models, and make the analysis presented inappropriate . Thus, it is dangerous to extrapolate the results found below to other models.

A summary description of the variables used here is given in Table 2.1. In most of the models it will be assumed that true cognitive development at the time of the pretest (X_1^*) is not available for statistical analysis; the pretest score (X_1) is assumed to be a fallible but unbiased measure of cognitive development for all children. Thus, situations where the cognitive measure is culturally biased and where the cognitive measure is biased for extremely able or disadvantaged children are not considered in this chapter.[1] All of the models in this chapter use a dummy variable (Z) for experimental status; it is assumed that only one experimental level is offered and the children either participate in Head Start or belong to the control group. This particular assumption is, however, not crucial to the analysis; it is used largely for convenience and because in the empirical analysis Head Start is considered as a discrete treatment. For an example of an analysis

Table 2.1

Description of Variables Used in the Theoretical Models

Variable	Description
X_1^*	True cognitive development at the time of the pretest
X_1	Measured cognitive development at the time of the pretest; i.e., pretest score
X_2	Socioeconomic status
Z	Dummy variable for treatment defined as $Z = \begin{cases} 1 \text{ if child received treatment, i.e., was in Head Start} \\ 0 \text{ if child did not receive treatment, i.e., in control group} \end{cases}$
u	Disturbance term associated with the pretest
v	Disturbance term associated with the posttest
Y	Measured cognitive development at the time of the posttest

of the effects of Head Start where the treatment is continuous the reader is referred to Watts and Horner (1968).

The reader may wonder why there is no variable Y^* analogous to X_1^* in Table 2.1. The reason is that errors in the measurement of the dependent variable do not cause problems in obtaining unbiased estimates of the regression coefficients as errors in measuring the independent variables often do. The proof of this is quite simple, and can be found in Kmenta (1971, p.320).

2. A Model with No Errors in the Independent Variables

The first model to be considered is one in which all of the independent variables are known and measured without error. Even though economists rarely have the opportunity to use such "ideal" data, a great deal of the empirical work done by them assumes implicitly that there are no errors in the variables. Formally, this model may be specified as:

$$Y = \beta_0 + \beta_1 X_1^* + \beta_Z Z + v \tag{1}$$

$$E(X_1^*) = \mu,\ E(v) = 0,\ \mathrm{Cov}(v, X_1^*) = \mathrm{Cov}(v, Z) = 0 \tag{2}$$

When we compute the population regression of Y on X_1^* and Z by least squares, we obtain the correct values for β_0, β_1, and β_Z.

For example, if we define $\tilde{\beta}_Z$ as the value of the regression coefficient for Z obtained from the population regression, the normal equations for the model are:

$$\mathrm{Var}(X_1^*)\tilde{\beta}_1 + \mathrm{Cov}(X_1^*,Z)\tilde{\beta}_Z = \mathrm{Cov}(X_1^*,Y) \qquad (3)$$

$$\mathrm{Cov}(X_1^*,Z)\tilde{\beta}_1 + \mathrm{Var}(Z)\tilde{\beta}_Z = \mathrm{Cov}(Z,Y) \qquad (4)$$

Solving equations (3) and (4) for $\tilde{\beta}_Z$ we find:

$$\tilde{\beta}_Z = \frac{\mathrm{Cov}(Y,Z)\mathrm{Var}(X_1^*) - \mathrm{Cov}(Y,X_1^*)\mathrm{Cov}(X_1^*,Z)}{\mathrm{Var}(X_1^*)\mathrm{Var}(Z) - \mathrm{Cov}(X_1^*,Z)\mathrm{Cov}(X_1^*,Z)} \qquad (5)$$

$$= \frac{[\beta_1\mathrm{Cov}(X_1^*,Z) + \beta_Z\mathrm{Var}(Z)]\mathrm{Var}(X_1^*) - [\beta_1\mathrm{Var}(X_1^*) + \beta_Z\mathrm{Cov}(X_1^*,Z)]\mathrm{Cov}(X_1^*,Z)}{\mathrm{Var}(X_1^*)\mathrm{Var}(Z) - \mathrm{Cov}(X_1^*,Z)\mathrm{Cov}(X_1^*,Z)}$$

$$= \beta_Z$$

There are several insights that can be gained from this simple model. First, we note that even when there are no measurement problems it is still essential to employ a control group. If we simply regressed posttest scores on pretest cognitive development for experimental children only, we would be unable to differentiate between normal cognitive growth and treatment effects; this problem would be especially great for programs in which the treatment is of long duration. In equation (1) β_0 measures additive growth (or decay) that is common to all children between the pretest and posttest, and β_1 measures growth that is proportional to the child's original level of development. Thus, β_0 and β_1 measure the "before and after" changes in cognitive development and β_Z measures the "with and without" differences; it is necessary to employ a control group to distinguish between the changes in cognitive development due to the treatment and those due to maturation.

We can also note that if the model presented in equations (1) and (2) is valid the only reason for assigning children to the experimental and control groups randomly is to increase the efficiency of the estimates. When randomization or matching is used then $E(X_1^*|Z = 1) = E(X_1^*|Z = 0)$ and $\mathrm{Cov}(X_1^*,Z) = 0$.

In principle there is no reason to expect programs such as Head Start to be equally effective for all types of children. The programs could be designed so that only children with a very low initial level of cognitive development will gain anything, or conversely the material presented to the children could be so difficult that only the most advantaged children gain anything. Herzog et al. (1972) have found "the less they have, the less they learn" to be true for their preschool programs, but it is impossible to separate the effects of the program difficulty and the learning ability of the children. The suggested interaction effect can be captured by modifying (1) to yield

$$Y = \beta_0 + \beta_1 X_1^* + \beta_Z Z + \beta_3 X_1^* Z + v . \qquad (6)$$

If β_3 is positive then Herzog's finding that the less they have the less they learn would be verified. Note that if $\beta_3 \neq 0$ it would be important to know the level of the child's pre-enrollment cognitive development in order to predict the benefits of Head Start.

3. The Campbell-Erlebacher Two-Population Model[2]

The first errors in variables model to be considered is one developed by Campbell and Erlebacher (1970). Their model demonstrates that if the children in the experimental and control groups are selected from two different populations with the control population having a higher initial level of cognitive development, regression analysis can produce a spurious negative treatment effect. Although a computer simulation rather than a formal model was used in their paper, it is not difficult to construct the general model that Campbell and Erlebacher deal with implicitly:

Experimental Group		Control Group	
(7E)	$X_1 = X_1^* + u$	(7C)	$X_1 = X_1^* + u$
(8E)	$Y = X_1^* + v$	(8C)	$Y = X_1^* + v$
(9E)	$X_1^* \sim N(\mu_E, \sigma_*^2)$	(9C)	$X_1^* \sim N(\mu_C, \sigma_*^2)$
(10E)	$u \sim N(0, \sigma^2)$	(10C)	$u \sim N(0, \sigma^2)$
(11E)	$v \sim N(0, \sigma^2)$	(11C)	$v \sim N(0, \sigma^2)$
(12E)	$\mathrm{Cov}(u, X_1^*) = \mathrm{Cov}(v, X_1^*) = \mathrm{Cov}(u, v) = 0$	(12C)	$\mathrm{Cov}(u, X_1^*) = \mathrm{Cov}(v, X_1^*) = \mathrm{Cov}(u, v) = 0$

Campbell and Erlebacher assume that the control population is more able than the experimental population so that $\mu_C > \mu_E$. In the above model the pretest and posttest scores, X_1 and Y, represent unbiased but fallible measures of cognitive development. The assumption that u and v are uncorrelated implies that if a child scores higher than his true level of cognitive development on the pretest we have no a priori knowledge concerning whether he will score higher or lower than his true level on the posttest. The assumption that the slope and constant are the same in (8E) and (8C) indicates that we are assuming that the treatment has no effect; this assumption is made only for convenience. The regression of posttest on pretest for the experimental group is:

$$E(Y|X_1) = E(Y) - \frac{\mathrm{Cov}(X_1,Y)}{\mathrm{Var}(X_1)} \cdot E(X_1) + \frac{\mathrm{Cov}(X_1,Y)}{\mathrm{Var}(X_1)} \cdot X_1 \qquad (13E)$$

$$= \mu_E - \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*)+\mathrm{Var}(u)} \cdot \mu_E + \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*)+\mathrm{Var}(u)} \cdot X_1$$

$$= (1-P)\mu_E + PX_1$$

where $$P = \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*)+\mathrm{Var}(u)} .$$

Since $0 \leq P \leq 1$, the slope of the regression line is attenuated due to the error in measurement.[3]

Because the only difference between the experimental and control groups is in their means on X_1^*, X_1, and Y, the regression equation for the control population can be written as

$$E(Y|X_1) = (1-P)\mu_C + PX_1 \ . \tag{13C}$$

The treatment variable Z is defined as

$$Z = \begin{cases} 1 \text{ if a subject had the treatment} \\ 0 \text{ if a subject did not have the treatment.} \end{cases} \tag{14}$$

In the Campbell-Erlebacher model Z corresponds exactly to population membership:

$$Z = \begin{cases} 1 \text{ if a subject was in the "lower" population} \\ 0 \text{ if a subject was in the "higher" population.} \end{cases} \tag{15}$$

Then (13E) and (13C) can be rewritten as

$$E(Y|X_1, Z=0) = (1-P)\mu_C + PX_1 \tag{16}$$

$$E(Y|X_1, Z=1) = (1-P)\mu_E + PX_1 \tag{17}$$

Because Z only takes on the values of 0 and 1, equations (16) and (17) can be combined to give this result:

$$E(Y|X_1, Z) = (1-P)\mu_C + PX_1 + (1-P)(\mu_E - \mu_C)Z. \tag{18}$$

Recalling that $\mu_E - \mu_C < 0$, we can plot the regression of Y on X_1 and Z:

Regression of Y on X_1 and Z for the Campbell-Erlebacher Model

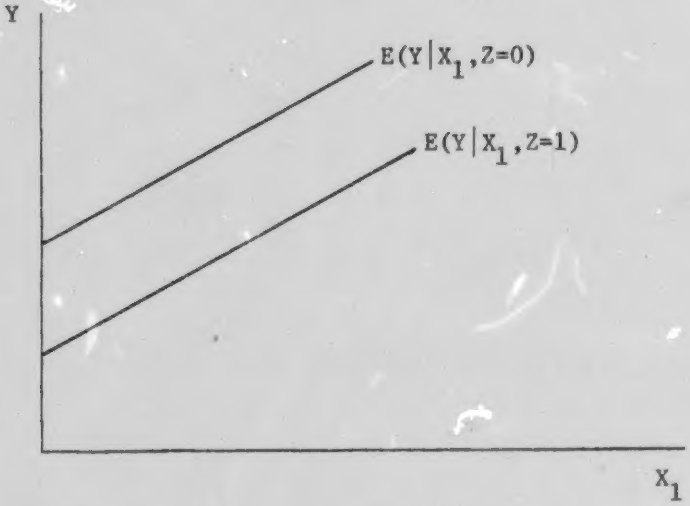

Figure 2.1

Clearly the coefficient of Z does not measure the effect of the treatment; it simply reflects the difference in population means. Group membership is serving as a proxy for cognitive development.

As it happens, the Campbell-Erlebacher model is structured so that Head Start could be evaluated without bias by regressing gain scores, $G = Y - X_1$, on treatment status Z. Campbell and Erlebacher (1970, p. 197) indicate that gain scores would avoid the bias problem, but they warn that "gain scores are in general such a treacherous quicksand, e.g., are so non comparable for high versus low scores within any particular sample, that one is reluctant to recommend them for any purpose." This warning has been given by other researchers such as Lord (1963), Bohrnstedt (1969), and Cronbach and Furly (1970). The use of gain scores as the dependent variable produces an unbiased measure of treatment effect only under special conditions. If equations (8E) and (8C) are modified to allow for proportional growth ($Y = aX_1^* + v$, $a \neq 1$) then gain scores would not lead to an unbiased measure of treatment effect. Similarly, gain scores are inappropriate in the one-population model discussed below when $\beta_1 \neq 1$. If the model is further complicated by the inclusion of heteroscedasticity or interaction effects, gain scores are also inappropriate; Lord (1963) discusses these problems in detail.

4. A One-Population Model with Nonrandom Selection

It is possible to object to the Campbell-Erlebacher model on the grounds that cognitive development is not necessarily distributed in two separate normal distributions but rather in one larger population. Differences in the preprogram levels of development between the experimental and control groups can still result from nonrandom selection procedures. The model developed below modifies the Campbell-Erlebacher model so that there is only one population, and is more general because it allows for linear and proportional growth. It should be noted that the model presented in this section is a general one for regression when there are errors of measurement for one regressor: we are simply applying the model to a Head Start evaluation.

We suppose the basic equation of the model to be:

$$Y = \beta_0 + \beta_1 X_1^* + \beta_Z Z + v \tag{19}$$

where the variables are defined as in Table 2.1. We further assume that X_1^* is unavailable for the evaluation, but that we do have the pretest score available:

$$X_1 = X_1^* + u \tag{20}$$

where u is independent of v, X_1^*, and Z. It is also assumed that v is independent of X_1^* and Z; more formally this can be stated as:

$$\begin{aligned}\mathrm{Cov}(u,Z) &= \mathrm{Cov}(u,v) = \mathrm{Cov}(u,X_1^*) = \mathrm{Cov}(v,X_1^*) = \mathrm{Cov}(v,Z)\\ &= 0. \end{aligned}\tag{21}$$

In addition, we shall assume that u, v, and X_1^* all have normal distributions. We are interested in determining if the regression coefficient of Z will be the same when we run the linear regression of Y on X_1 and Z rather than Y on X_1^* and Z.[4] Thus when we determine

$$E(Y|X_1,Z) = \alpha_0 + \alpha_1 X_1 + \alpha_Z Z \tag{22}$$

will $\alpha_1 = \beta_1$ and $\alpha_Z = \beta_Z$? First, let us make the following definitions:

$$\sigma_{11} = \mathrm{Var}(X_1^*),\ \sigma_{ZZ} = \mathrm{Var}(Z),\ \sigma_{1Z} = \mathrm{Cov}(X_1^*, Z),\ \sigma_{uu} = \mathrm{Var}(u)$$

$$P = \frac{\sigma_{11}}{\sigma_{11}+\sigma_{uu}}\ ,\ r^2 = \frac{(\sigma_{1Z})^2}{\sigma_{11}\sigma_{ZZ}}\quad . \tag{23}$$

Note that the parameter P in this model is defined as the ratio of the variance of true cognitive development to the variance of measured development for the entire population, whereas in the Campbell-Erlebacher model, P was the ratio of the within-group variances. The parameter r^2 is the squared coefficient of correlation between X_1^* and Z; we know that $0 \le r^2 \le 1$.

We can now use the normal equations to solve for α_1 and α_Z in terms of β_1 and β_Z:

$$\alpha_1 = \frac{\mathrm{Cov}(Y,X_1)\cdot\mathrm{Var}(Z) - \mathrm{Cov}(Y,Z)\cdot\mathrm{Cov}(X_1,Z)}{\mathrm{Var}(X_1)\ \cdot\ \mathrm{Var}(Z) - \mathrm{Cov}(X_1,Z)\cdot\mathrm{Cov}(X_1,Z)} \tag{24}$$

$$= \frac{(\beta_1\sigma_{11} + \beta_Z\sigma_{1Z})\sigma_{ZZ} - (\beta_1\sigma_{1Z} + \beta_Z\sigma_{ZZ})\sigma_{1Z}}{(\sigma_{11} + \mathrm{Var}(u))\sigma_{ZZ} - \sigma_{1Z}\sigma_{1Z}}$$

$$= \frac{P(1-r^2)}{1-Pr^2}\ \cdot\ \beta_1$$

$$\alpha_Z = \frac{\mathrm{Var}(X_1) \cdot \mathrm{Cov}(Z,Y) - \mathrm{Cov}(X_1,Z) \cdot \mathrm{Cov}(X_1,Y)}{\mathrm{Var}(X_1) \cdot \mathrm{Var}(Z) - \mathrm{Cov}(X_1,Z) \cdot \mathrm{Cov}(X_1,Z)} \tag{25}$$

$$= \frac{\frac{\sigma_{11}}{P}(\beta_1\sigma_{1Z} + \beta_Z\sigma_{ZZ}) - \sigma_{1Z}(\beta_1\sigma_{11} + \beta_Z\sigma_{1Z})}{\frac{\sigma_{11}}{P} \cdot \sigma_{ZZ} - \sigma_{1Z}\sigma_{1Z}}$$

$$= \beta_Z + \frac{\sigma_{1Z}(1-P)}{\sigma_{ZZ}(1-Pr^2)} \cdot \beta_1$$

Thus we find that in general $\alpha_1 \neq \beta_1$ and $\alpha_Z \neq \beta_Z$.

For an evaluation of Head Start we are particularly interested in knowing when $\alpha_Z = \beta_Z$. One case in which $\alpha_Z = \beta_Z$ is when $P = 1$; for this to be true, however, Var(u) must equal zero, so X_1^* equals X_1 for all observations. Clearly, $P = 1$ implies that we have no measurement error and the model reduces to the model discussed in section 2, where we could directly measure β_Z. Another case in which there would be no bias is when $\beta_1 = 0$; i.e., the variable measured with error does not belong in the regression. This case is theoretically untenable because it implies that the child's level of cognitive development in the second period has no relationship to his cognitive development in the first period. A more interesting case is that $\sigma_{1Z} = 0$

suffices for $\alpha_Z = \beta_Z$. But $\sigma_{1Z} = 0$ is equivalent to $E(X_1^*|Z) = E(X_1^*)$, which says that the mean of initial cognitive development is the same in the experimental and control groups. Random selection or matching on pretests would lead to this result and would therefore eliminate bias. Note that if $\sigma_{1Z} < 0$ (the experimental group initially has lower cognitive development) then $\alpha_Z < \beta_Z$ and the bias would be to underestimate treatment effect. This corresponds to the Campbell-Erlebacher result where a less able experimental group led to an underestimate of the treatment effect.

There are several selection procedures that are compatible with the model described above which would lead to an underestimate of treatment effect. One such procedure is that the Head Start officials knew the values of X_1^* and put the upper half of the population in the control group and selected the bottom half who "needed" the treatment more for the experimental group; this selection procedure is discussed by Goldberger (1972a) and Barnow (1972). This type of selection procedure is interesting theoretically but is of little practical value because the program evaluators would then also have the observations on X_1^* available. Selection on the basis of pretests (X_1) is a possibility, but violates assumption (21) because $\text{Cov}(u,Z) \neq 0$; this model will be considered in section 6. There is a selection procedure that would not violate the assumptions of the model and is still tenable. This is for group selection to be made on the basis of one pretest and for another pretest to be used for the evaluation. Suppose, for example, that the Head Start officials interviewed all of the eligible families and ranked them on the "need" of their children to participate in Head Start. (This "pretest" would not need to be a formal test but could be done by gathering information on the socioeconomic status of the families and observing the children to obtain a score.) The Head Start officials could then assign the half of the population whom they felt would benefit most from Head Start--e.g., the low SES ones--to the experimental group, and the other half of the children would be assigned to the control group. Both groups would then be given a formal pretest for use later in the evaluation of Head Start. If we call the screening pretest $\tilde{X}_1$, we would add the following equations to the model:

$$\tilde{X}_1 = a_0 + a_1 X_1^* + w \tag{26}$$

$$\mathrm{Cov}(u,w) = \mathrm{Cov}(w,X_1^*) = \mathrm{Cov}(v,w) = 0. \tag{27}$$

If $\tilde{X}_1 \leq \tilde{\bar{X}}_1$ the child would be assigned to the experimental group ($Z = 1$), and if $\tilde{X}_1 > \tilde{\bar{X}}_1$ the child would be assigned to the control group ($Z = 0$). These additions to the model would not violate any of the previous assumptions, including the assumption that $\mathrm{Cov}(u,Z) = 0$. This selection procedure, which is probably more realistic, we shall call "selecting on one pretest and controlling on another," and it leads to the same bias direction as in the Campbell-Erlebacher model.

5. A One-Population Model with Interaction

When the model described in the previous section is expanded to allow for an interaction effect for treatment and initial level of cognitive development, it becomes considerably more difficult to analyze. The basic equation of the model is:

$$Y = \beta_0 + \beta_1 X_1^* + \beta_Z Z + \beta_3 X_1^* Z + v \ . \tag{28}$$

We now define the variable X_3^* to represent the product of Z and X_1^*:

$$X_3^* = X_1^* Z \ . \tag{29}$$

Again, it is assumed the X_1^* is unobservable and that only a fallible measure of X_1^* is available:

$$X_1 = X_1^* + u \tag{30}$$

$$\begin{aligned} X_3 &= X_1 Z \\ &= X_1^* Z + uZ \end{aligned} \tag{31}$$

$$\mathrm{Cov}(u,v) = \mathrm{Cov}(u,X_1^*) = \mathrm{Cov}(v,X_1^*) = \mathrm{Cov}(u,Z) = \mathrm{Cov}(v,Z) = 0. \tag{32}$$

We are thus unable to directly estimate (28), but by using X_1 and X_3 we can calculate the linear regression of Y on X_1, Z, and X_3:

$$E(Y|X_1,Z,X_3) = \alpha_0 + \alpha_1 X_1 + \alpha_Z Z + \alpha_3 X_3 \tag{33}$$

The normal equations for α_1, α_Z, and α_3 are:

$$\mathrm{Var}(X_1)\alpha_1 + \mathrm{Cov}(X_1,Z)\alpha_Z + \mathrm{Cov}(X_1,X_3)\alpha_3 = \mathrm{Cov}(X_1,Y) \tag{34}$$

$$\mathrm{Cov}(X_1,Z)\alpha_1 + \mathrm{Var}(Z)\alpha_Z + \mathrm{Cov}(Z,X_3)\alpha_3 = \mathrm{Cov}(Z,Y) \tag{35}$$

$$\mathrm{Cov}(X_1,X_3)\alpha_1 + \mathrm{Cov}(Z,X_3)\alpha_Z + \mathrm{Var}(X_3)\alpha_3 = \mathrm{Cov}(X_3,Y) \tag{36}$$

If we solve equations (34), (35), and (36) we find that in general the α coefficients will not be equal to the corresponding β coefficients. In the previous model where it was assumed there was only an additive treatment effect it was demonstrated that random selection would produce an unbiased coefficient for treatment effect. In the present model α_Z, the coefficient for additive treatment effect is similarly unbiased, but we find that

$$\alpha_3 = \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*)+\mathrm{Var}(u)} \cdot \beta_3 \leq \beta_3 \tag{37}$$

when one-half of the population is randomly assigned to each group.[5] Thus the coefficient of X_3 will be attenuated when there is measurement error.

Goldberger (1972b) has examined the interaction model where the selection of group membership is determined entirely by X_1^* with the upper half of the population placed in the control group and the lower half in the experimental group. He finds that both α_Z and α_3 are biased, and that α_3 is more attenuated than when group membership is determined randomly.

It should be noted that selection on the basis of pretests violates the assumptions of the model; this selection procedure will be considered in the following section.

6. Selection on the Basis of Pretest Scores[6]

The model presented in section 4 is valid only when the error associated with the pretest (u) is uncorrelated with treatment status (Z); more formally, the model assumed that Cov(u,Z) = 0. It was demonstrated that this assumption is consistent with certain selection procedures--random selection, selection on the basis of true scores, and selection on one pretest and control on another. We now consider the case where group assignment is based on pretest score and where the pretest score is available as an independent variable for statistical analysis. Since pretest scores are correlated (although not perfectly) with true scores, when we select on the former we are also selecting, in a sense, on the true score. Lord and Novick (1968, p. 141) refer to these two methods of selection as explicit selection (selection on the basis of X_1^*) and incidental selection (selection on X_1, the fallible measure of X_1^*). Because we have discovered that selection on the basis of X_1^*, true ability, leads to bias in the coefficient of Z, we might expect that selection on the basis of pretest scores would lead to the same result. The analysis below demonstrates that this belief is incorrect; selection on the basis of pretests does not lead to biased treatment coefficients.

The basic equations of this model are:

$$Y = \beta_0 + \beta_1 X_1^* + \beta_Z Z + v \qquad (38)$$

$$X_1 = X_1^* + u \qquad (39)$$

$$X_1^* \sim N(\mu, \sigma_*^2) \qquad (40)$$

$$u \sim N(0, \sigma^2) \qquad (41)$$

$$v \sim N(0, \sigma^2) \qquad (42)$$

$$\text{Cov}(u,v) = \text{Cov}(u,X_1^*) = \text{Cov}(v,X_1^*) = \text{Cov}(v,Z) = 0 \qquad (43)$$

Because we assume that X_1^* is unobservable, we want to determine if the regression

$$E(Y|X_1) = \alpha_0 + \alpha_1 X_1 + \alpha_Z Z \qquad (44)$$

will lead to $\alpha_Z = \beta_Z$. The normal equations for the population are

$$\text{Var}(X_1)\alpha_1 + \text{Cov}(X_1,Z)\alpha_Z = \text{Cov}(X_1,Y) \qquad (45)$$

$$\text{Cov}(X_1,Z)\alpha_1 + \text{Var}(Z)\alpha_Z = \text{Cov}(Z,Y) \qquad (46)$$

To solve for α_Z we must determine the values of the variance and covariance terms in (45) and (46). For covenience we assume that after the children have been ranked on the basis of their pretests the upper half is assigned to the control group and the lower half is assigned to the experimental group. We have already demonstrated that $\mathrm{Var}(Z) = \frac{1}{4}$ and $\mathrm{Var}(X_1) = \sigma_*^2 + \sigma^2$ for this situation. To solve for the covariance terms we make substantial use of the theorems proved by Goldberger (1972a), but we omit the proofs. The first theorem we make use of is: If Z is a binary variable with $E(Z) = \frac{1}{2}$ and if w is some other variable, then

$$\mathrm{Cov}(Z,w) = -\tfrac{1}{4}[E(w|Z=0) - E(w|Z=1)]. \qquad (47)$$

To calculate $\mathrm{Cov}(Z,X_1)$ and $\mathrm{Cov}(Z,Y)$ we must calculate the conditional expectations of X_1 and Y for the control and experimental groups. Because X_1 is distributed normally and those in the upper half of the distribution have Z = 1 and those in the lower half have Z = 0, finding the conditional expectation of X_1 given Z reduces to the problem of determining the conditional expectation of X_1 given the half of the normal distribution X_1 is in. Goldberger (1972a, p. 9) has shown that this conditional expectation can be written as:

$$E(w|Z) = \tilde{\mu} + (1-2Z)\,\tilde{\sigma}\,\sqrt{2/\pi} \qquad (48)$$

where $\tilde{\sigma}$ and $\tilde{\mu}$ are the standard deviation and mean, respectively, of the normally distributed variable w. Using (47) and (48) we can determine $\mathrm{Cov}(Z,X_1)$:

$$\mathrm{Cov}(Z,X_1) = -\tfrac{1}{4}\,[\mu + \sqrt{2(\sigma_*^2 + \sigma^2)/\pi} \tag{49}$$

$$- \mu + \sqrt{2(\sigma_*^2 + \sigma^2)/\pi}\ ]$$

$$= -\sqrt{(\sigma_*^2 + \sigma^2)/2\pi}$$

The value of $E(Y|Z)$ is more difficult to determine. We begin by calculating

$$E(u|X_1) = \frac{\mathrm{Cov}(u,X_1)}{\mathrm{Var}(X_1)} \cdot X_1 - \frac{\mathrm{Cov}(u,X_1)}{\mathrm{Var}(X_1)} \cdot \mu \tag{50}$$

$$= \frac{\sigma^2}{\sigma_*^2 + \sigma^2}\,(X_1 - \mu)$$

from the population regression of u on X_1. We then note from Goldberger (1972a, p.15) that because Z is an exact function of X_1 we can write

$$E(u|Z) = E[(E(u|X_1))|Z] \tag{51}$$

$$= E[(\frac{\sigma^2}{\sigma_*^2 + \sigma^2}\ (X_1 - \mu))|Z]$$

$$= \frac{\sigma^2}{\sigma_*^2 + \sigma^2}\,[E(X_1|Z) - \mu]$$

$$= \frac{\sigma^2}{\sigma_*^2 + \sigma^2}\,(1 - 2Z)\ \sqrt{2(\sigma_*^2 + \sigma^2)/\pi}$$

Now that $E(u|Z)$ and $E(X_1|Z)$ are known, we can calculate $E(X_1^*|Z)$, $E(Y|Z)$, and then $\mathrm{Cov}(Z,Y)$.

$$E(X_1^*|Z) = E(X_1|Z) - E(u|Z) \tag{52}$$

$$= \mu + (1-2Z)\sqrt{2(\sigma_*^2+\sigma^2)/\pi}$$

$$- \frac{\sigma^2}{\sigma_*^2+\sigma^2} \, (1-2Z)\sqrt{2(\sigma_*^2+\sigma^2)/\pi}$$

$$= \mu + \frac{\sigma_*^2}{\sigma_*^2+\sigma^2} \, (1-2Z) \, \sqrt{2(\sigma_*^2+\sigma^2)/\pi}$$

$$E(Y|Z) = \beta_0 + \beta_1 E(X_1^*|Z) + \beta_Z Z + E(v|Z) \tag{53}$$

$$= \beta_0 + \beta_1\left(\mu + \frac{\sigma_*^2}{\sigma_*^2+\sigma^2} \, (1-2Z)\sqrt{2(\sigma_*^2+\sigma^2)/\pi)}\right) + \beta_Z Z$$

$$\mathrm{Cov}(Y,Z) = -\tfrac{1}{4}(E(Y|Z=0) - E(Y|Z=1)) \tag{54}$$

$$= -\tfrac{1}{4}\left(\frac{2\beta_1\sigma_*^2}{\sigma_*^2+\sigma^2} \cdot \sqrt{2(\sigma_*^2+\sigma^2)|\pi} - \beta_Z\right)$$

$$= \tfrac{1}{4}\beta_Z - \tfrac{1}{2} \cdot \frac{\beta_1\sigma_*^2}{\sigma_*^2+\sigma^2} \cdot \sqrt{2(\sigma_*^2+\sigma^2)/\pi}$$

The only covariance term that remains to be determined is $\mathrm{Cov}(X_1,Y)$:

$$\begin{aligned}\mathrm{Cov}(X_1,Y) &= E(X_1Y) - E(X_1)\cdot E(Y) \qquad (55)\\ &= E\Big((X_1^\star + u)\ (\beta_0 + \beta_1X_1^\star + \beta_Z Z + v)\Big)\\ &\quad - \mu(\beta_0 + \beta_1\mu + \tfrac12\beta_Z)\\ &= \beta_1(E(X_1^{\star 2}) - \mu^2) + \beta_Z(E(X_1Z) - \tfrac12\mu)\\ &= \beta_1\mathrm{Var}(X_1^\star) + \beta_Z\mathrm{Cov}(X_1,Z)\\ &= \beta_1\sigma_\star^2 - \beta_Z\sqrt{(\sigma_\star^2+\sigma^2)/2\pi}\end{aligned}$$

We can now solve the normal equations in (45) and (46) by substituting the variances and covariances found in (47), (49), (54), and (55) into them:

$$\alpha_Z = \frac{\mathrm{Var}(X_1)\cdot\mathrm{Cov}(Z,Y) - \mathrm{Cov}(X_1,Z)\cdot\mathrm{Cov}(X_1,Y)}{\mathrm{Var}(X_1)\cdot\mathrm{Var}(Z) - \mathrm{Cov}(X_1,Z)\cdot\mathrm{Cov}(X_1,Z)} \qquad (56)$$

$$= \frac{(\sigma_\star^2+\sigma^2)\left(\tfrac14\beta_Z - \tfrac12\beta_1\cdot\dfrac{\sigma_\star^2}{\sigma_\star^2+\sigma^2}\cdot\sqrt{2(\sigma_\star^2+\sigma^2/\pi}\right)}{\tfrac14(\sigma_\star^2+\sigma^2) - (\sigma_\star^2+\sigma^2)/2\pi}$$

$$\frac{+\ \sqrt{(\sigma_\star^2+\sigma^2)/2\pi}\left(\beta_1\sigma_\star^2 - \beta_Z\sqrt{(\sigma_\star^2+\sigma^2)/2\pi}\right)}{\tfrac14(\sigma_\star^2+\sigma^2) - (\sigma_\star^2+\sigma^2)/2\pi}$$

$$= \frac{\beta_Z[\tfrac14(\sigma_\star^2+\sigma^2) - (\sigma_\star^2+\sigma^2)/2\pi] + \beta_1\left[(\sigma_\star^2-\sigma_\star^2)\sqrt{(\sigma_\star^2+\sigma^2)/2\pi}\ \right]}{\tfrac14(\sigma_\star^2+\sigma^2) - (\sigma_\star^2+\sigma^2)/2\pi}$$

$$= \beta_Z\ .$$

$$\alpha_1 = \frac{\mathrm{Cov}(X_1,Y) \cdot \mathrm{Var}(Z) - \mathrm{Cov}(X_1,Z) \cdot \mathrm{Cov}(Z,Y)}{\mathrm{Var}(X_1) \cdot \mathrm{Var}(Z) - \mathrm{Cov}(X_1,Z) \cdot \mathrm{Cov}(X_1,Z)} \tag{57}$$

$$= \frac{\tfrac{1}{4}\left(\beta_1\sigma_*^2 - \beta_Z\sqrt{(\sigma_*^2 + \sigma^2)/2\Pi}\right)}{\tfrac{1}{4}(\sigma_*^2 + \sigma^2) - (\sigma_*^2 + \sigma^2)/2\Pi}$$

$$+ \frac{\sqrt{(\sigma_*^2 + \sigma^2)/2\Pi}\ \left[\tfrac{1}{4}\beta_Z - \tfrac{1}{2}\frac{\beta_1\sigma_*^2}{\sigma_*^2+\sigma^2}\left(\sqrt{2(\sigma_*^2+ \sigma^2)/\Pi}\right)\right]}{\tfrac{1}{4}(\sigma_*^2 + \sigma^2) - (\sigma_*^2 + \sigma^2)/2\Pi}$$

$$= \frac{\sigma_*^2}{\sigma_*^2+\sigma^2} \cdot \beta_1$$

Thus we find that explicit selection on X_1, the fallible indicator of X_1^*, does not lead to a biased coefficient for treatment effect. We do not find bias because Z is determined by X_1; the dummy variable for the experimental group provides no information about the child's preprogram level of cognitive development that is not already given by X_1 itself. Hence, the coefficient of Z measures only the treatment effect. This selection procedure counters the claims of Campbell and Erlebacher since the treatment is made available to the most needy children, yet the coefficient for Head Start remains unbiased.

Campbell (1969) refers to this type of experiment as a "regression discontinuity design" and recommends its use when the experimenter is unwilling to use random selection procedures. Selection on the basis of pretests is, incidentally, attractive because it permits those who need the program most to receive the benefits. The major disadvantage of using selection on pretest rather than random selection is that random selection is more efficient. The loss of efficiency results from the introduction of multicollinearity between X_1 and Z when selection is based on pretests. Goldberger (1972a) shows that random selection is approximately 2.75 times more efficient than selection on pretests; a random sample of 100 is as good as a sample of 275 where selection is on the basis of pretests.

We shall not formally present the pretest-selection model with interaction effects; this model is discussed in detail by Goldberger (1972b). The results are similar to our findings for the present model: selection on the basis of pretests produces the same coefficients as random selection for the population regression but is less efficient than random selection. In section 4 of this chapter it was demonstrated that when there is an interaction effect, random selection leads to an unbiased coefficient for the additive effect but an attenuated coefficient for the interaction effect. When selection is based on pretest scores the coefficient for the additive effect is again unbiased, and the coefficient for the interaction effect is attenuated to the same degree as for random selection.

7. An Omitted Variable Model

All of the models presented in this chapter have assumed the existence and availability of a pretest to measure cognitive development before the commencement of the Head Start program. We now examine a model that is explicitly ex post facto--we assume that there are no measures of the initial level of cognitive development available for any of the children. The model to be considered is

$$Y = \beta_0 + \beta_1 X_1^* + \beta_2 X_2 + \beta_Z Z + v. \tag{58}$$

In this model it is useful to consider X_1^* as representative of traits determined in part by inheritance as well as environment, such as IQ. The variable X_2 which has not been included in the previous models represents socioeconomic status (SES) or home environment. We assume that X_1^* and X_2 have independent effects on the posttest score Y. Because SES is likely to be a multidimensional variable, the reader can view X_2 as a vector of variables rather than as a single variable; we use a single variable for SES solely to simplify the analysis of the model. The variable Z is again a dummy variable for experimental status and v is the disturbance term associated with the posttest score. We shall assume that X_1^*, X_2, and u have normal distributions and that v is independent of X_1^* and X_2. Because this model is for an ex post facto study, X_1^*, X_2, and Z cannot be measured at the time when the program begins; we assume that X_1^* is unobserved and that X_2 does not change during the experiment. If X_2 does change during the experiment and if the change is not simply a linear transformation on all observations then the model must be expanded to account for the measurement error in X_2.

If we had observations on Y, X_1^*, X_2, and Z we could regress Y on X_1^*, X_2, and Z to directly determine β_Z, the effect of Head Start on cognitive development. However, we assume that X_1^* is not observed and we must determine if the regression

$$E(Y|X_2,Z) = \alpha_0 + \alpha_2 X_2 + \alpha_Z Z \tag{59}$$

will yield $\alpha_Z = \beta_Z$. For convenience we use the following notation:

$\sigma_{11} = \mathrm{Var}(X_1^*)$ $\quad\sigma_{12} = \mathrm{Cov}(X_1^*,X_2)$ $\quad\sigma_{1Z} = \mathrm{Cov}(X_1^*,Z)$

$\sigma_{1Y} = \mathrm{Cov}(X_1^*,Y)$ $\quad\sigma_{22} = \mathrm{Var}(X_2)$ $\quad\sigma_{2Z} = \mathrm{Cov}(X_2,Z)$

$\sigma_{2Y} = \mathrm{Cov}(X_2,Y)$ $\quad\sigma_{ZZ} = \mathrm{Var}(Z)$ $\quad\sigma_{ZY} = \mathrm{Cov}(Z,Y)$

When we solve the normal equations for (59) in terms of the betas we find that:

$$\alpha_Z = \frac{\sigma_{22}\sigma_{ZY} - \sigma_{2Z}\sigma_{2Y}}{\sigma_{22}\sigma_{ZZ} - \sigma_{2Z}\sigma_{2Z}} \tag{60}$$

$$= \frac{\sigma_{22}(\sigma_{1Z}\beta_1 + \sigma_{2Z}\beta_2 + \sigma_{ZZ}\beta_Z) - \sigma_{2Z}(\sigma_{12}\beta_1 + \sigma_{22}\beta_2 + \sigma_{2Z}\beta_Z)}{\sigma_{22}\sigma_{ZZ} - \sigma_{2Z}\sigma_{2Z}}$$

$$= \beta_Z + \frac{\sigma_{1Z}\sigma_{22} - \sigma_{12}\sigma_{2Z}}{\sigma_{22}\sigma_{ZZ} - \sigma_{2Z}\sigma_{2Z}} \cdot \beta_1$$

This is the standard result when a relevant variable is omitted; the coefficients of the included variables are in general biased. It can be demonstrated that

$$\frac{\sigma_{1Z}\sigma_{22} - \sigma_{12}\sigma_{2Z}}{\sigma_{22}\sigma_{ZZ} - \sigma_{2Z}\sigma_{2Z}} = b_{1Z\cdot 2} \tag{61}$$

where $b_{1Z\cdot 2}$ is the partial regression coefficient of Z when X_1^* is regressed on Z and X_2 (see Kmenta (1971, pp. 392-395)). (This is purely a mechanical relationship and does not depend on a causal model where X_1^* is determined by Z and X_2.)

The extent and direction of the bias of α_Z will depend upon β_1 and the variances and covariances of X_1^*, X_2, and Z. There are several cases of interest where $\alpha_Z = \beta_Z$. If random selection is used to assign the children to the experimental and control groups, then $\sigma_{1Z} = \sigma_{2Z} = 0$ and there will be no bias in α_Z. Note that a matching procedure where the control group is matched to the experimental group on X_2 will make $\sigma_{2Z} = 0$ but this is not sufficient for $\alpha_Z = \beta_Z$. Since partial correlation coefficients always have the same signs as the partial regression coefficients, $r_{1Z\cdot2} = 0$ is a sufficient condition for $b_{1Z\cdot2} = 0$ and hence for $\alpha_Z = \beta_Z$. Thus if children at any given SES level are assigned randomly, with regard to X_1^*, to the experimental and control groups, regression analysis will produce no spurious treatment effect. To clarify this point, consider the following example. Suppose that children are stratified by SES into three groups--high, middle, and low. Further assume that the administrators of Head Start are primarily interested in helping disadvantaged children. Then assume that 90 percent of the low SES group are selected at random and assigned to the experimental group with the remaining 10 percent assigned to the control group. For the middle SES group we shall assume that 50 percent are selected at random to receive Head Start and the other 50 percent assigned to the control group. Finally we assume that 10 percent of the high SES group is selected at random and assigned to the Head Start group with the other 90 percent assigned to the control group. If this selection procedure is used we know that r_{1Z}, the simple correlation of initial cognitive development and treatment status, will not be zero because most of the more able children will be assigned to the control group (assuming that SES and cognitive development are positively correlated). Nevertheless, the random selection within each SES group assures us that

$r_{1Z.2}$ will be zero because within each SES group selection is random. Thus this selection procedure will not produce a spurious treatment effect.

8. Application of the Models to the Westinghouse Study

Because the Westinghouse study was carried out ex post facto with the control group selected after the experimental group had received Head Start, there are no pretests available, and it may appear that the models with pretests are not applicable. If several assumptions are made, however, we can interpret the SES information that is available to be similar to a pretest. The assumptions that must be made to view a composite index of SES (such as the Hollingshead Index used in the Westinghouse study) or a vector of SES variables as a measure of cognitive development prior to the Head Start experience of the experimental children are that: (1) the SES variable or vector of variables is a function of cognitive development, (2) we know how to specify the functional relationship between SES and cognitive development, and (3) exposure to Head Start does not affect the SES of a child. If we make these assumptions then we can use the SES information as a fallible measure of pretreatment cognitive development in a statistical analysis.

The two-population model, as defined by Campbell and Erlebacher, has the property of being irrefutable; the model assumes that the children in the experimental group were selected from a less able population than the children in the control group. Even if the experimental and control groups have the same pretest means because of a matching procedure, Campbell and Erlebacher argue that the groups will differ on their true score means and that the effects of Head Start will be underestimated. Thus Campbell and Erlebacher appear to reject all nonrandom selection procedures for evaluation of education programs.

Because the two-population model begs the question of selection procedure, the one-population model presented in section 4 offers more help for empirical analysis. For the one-population model discussed in section 4 we found that if pretreatment cognitive development is uncorrelated with treatment status, the coefficient for treatment effect, α_Z, will be unbiased when the posttest score is regressed on the pretest score and the Head Start variable. Although no pretest scores are available from the Westinghouse data, we have noted that the vector of socioeconomic and demographic variables can be used as a proxy for pretreatment cognitive development. To determine if the groups differ on this vector of variables discriminant analysis can be used. If the selection procedure used for assigning children to the Head Start and control groups was compatible with the assumptions of the model described in section 4 then discriminant analysis enables us to discover if regression analysis will lead to an unbiased estimate of the effect of Head Start.

There are, however, several problems that reduce the usefulness of discriminant analysis for determining the presence or absence of bias when we must discriminate on the basis of SES rather than on actual pretests. Using discriminant analysis it can be determined whether the two groups are significantly different on a set of variables, but there is no method of translating these differences in SES into the metrics used in the posttest scores. If variables that are not relevant to cognitive development are included in the analysis, the test of group differences will be irrelevant; if the control group had more children with blue eyes, for example, we would not want this fact to influence the test of whether or not the experimental and control groups differ on the basis of cognitive development. Thus we must be careful to include only relevant variables in the discriminant analysis. Discriminant analysis is also unable to account for offsetting differences between the two groups. It is possible for the control group to have higher family incomes and for the experimental group to have more highly educated parents; we would expect the first example to favor the control group and the second to favor the experimental group so that the net effects would tend to offset each other. Discriminant analysis will only test whether or not the groups are different in terms of the included variables but does not have the power to account for these offsetting differences.

The most important problem with the use of discriminant analysis is that there are several selection procedures that will not lead to biased coefficients for treatment even though the groups differ on their preprogram level of cognitive development. In this chapter the model where group membership is determined by pretest scores and the model where selection within SES class is random will yield unbiased treatment coefficients even though the groups differ on their preprogram level of cognitive development.

Thus there is no simple way to determine if an unbiased treatment coefficient can be obtained unless we know the selection procedure used. For an ex post facto study such as the Westinghouse study, where selection was carried out in a very decentralized manner with 104 units selecting the members of the experimental group, it is impossible to know which selection procedure or procedures were used. Thus, the bias issue cannot be resolved at this time for the Westinghouse data.

FOOTNOTES

[1] The issue of bias in the test instruments is especially acute for evaluation of programs such as Head Start that serve children from different cultures and that try to serve children with low levels of cognitive ability. If a test instrument has a systematic bias for children from minority groups (i.e., Mexican American children always score 10 points lower than white children of the same ability) then the groups can be analyzed separately. If the test used does not measure the skill or ability that it purports to or measures one that is not of interest to the evaluator, then one must exercise caution when interpreting the findings of the evaluation. In our empirical work we have attempted to resolve these problems by analyzing the effects of Head Start separately for each ethnic group and by omitting those children from our samples who scored very low on the tests.

[2] The treatment of this and the following models draws heavily on the work of Barnow (1972) and Goldberger (1972a) and (1972b).

[3] The parameter we have called P is analogous to what is often referred to as the reliability coefficient of a test. If the value of P is known for the population under consideration the magnitude of the attenuation of the slope can be determined.

[4] Goldberger (1972a) has shown that the true regression of Y on X_1 is not linear when selection is made on the basis of X_1^*. Because empirical work is generally run using linear approximations, using a linear regression reflects what will happen in actual experiments. In qualitative terms, the spurious treatment effect retains the same direction of bias, but since the within-group regressions are no longer parallel the treatment effect calculated in the nonlinear regression will be a function of X_1.

[5] The simplest way to get this result is to transform all variables by using deviations from their means. We can then compute the appropriate variances and covariances when selection is random (thus assuming that X_1 and X_3 are uncorrelated with Z):

$\mathrm{Var}(X_1) = \mathrm{Var}(X_1^*) + \mathrm{Var}(u)$ $\quad \mathrm{Cov}(X_1,Z) = 0$ $\quad \mathrm{Cov}(X_1,X_3) = 0$

$\mathrm{Cov}(X_1,Y) = \beta_1 \mathrm{Var}(X_1^*)$ $\quad \mathrm{Var}(Z) = 1/4$ $\quad \mathrm{Cov}(Z,X_3) = 0$

$\mathrm{Cov}(Z,Y) = 1/4\ \beta_Z$ $\quad \mathrm{Var}(X_3) = 1/4[\mathrm{Var}(X_1^*) + \mathrm{Var}(u)]$ $\quad \mathrm{Cov}(X_3,Y) = 1/4\ \beta_3 \mathrm{Var}(X_1^*)$

These values can then be used to solve the normal equations in (34), (35), and (36) to get

$$\alpha_1 = \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*) + \mathrm{Var}(u)} \cdot \beta_1$$

$$\alpha_3 = \frac{\mathrm{Var}(X_1^*)}{\mathrm{Var}(X_1^*) + \mathrm{Var}(u)} \cdot \beta_3$$

$$\alpha_Z = \beta_Z$$

[6] This section is based on the presentation in Goldberger (1972b).

Chapter 3

Review of the Westinghouse Study

1. Introduction and Background

In this chapter a brief review of the Westinghouse study will be presented. We will discuss the rationale for the study, the methods that were used in gathering the data, the techniques that were used to analyze the data, the statistical findings of the analysis, and the interpretations and conclusions that were made by the Westinghouse staff. In Chapter 4 criticisms of the study that have appeared in journals and books will be both presented and evaluated. Additional comments on how the Westinghouse evaluation could have been improved will be offered. Thus, this chapter will not dwell on the shortcomings of the Westinghouse evaluation but will attempt only to summarize the actual 1969 study.

The Westinghouse study was carried out pursuant to a contract awarded by the Office of Economic Opportunity (OEO) to the Westinghouse Learning Corporation in collaboration with Ohio University. The Westinghouse researchers responded to a Request for Proposals (RFP) issued by OEO on April 22, 1968, and began their research in June 1968. The RFP was very specific concerning what the scope of the Head Start evaluation was to be. Cicirelli et al. (1969, p. 14) state that the RFP required the research to include:

> (1) A specific focus on the cognitive and affective development of Head Start enrollees;
>
> (2) An investigation and assessment of the residual or long-term effect of Head Start on these particular dimensions through

> examination of former Head Start enrollees in the early grades of primary school;
>
> (3) An evaluation of such Head Start impact, on the average, across the country, as well as by certain selected subgroups of local centers; and
>
> (4) A careful measurement and comparative assessment of the impact on children's intellectual and sociopersonal development of the two main types of programs: summer and full year.

(For the remainder of this dissertation, Cicirelli et al. (1969) is referred to as the Westinghouse study, and unless otherwise designated, page references refer to the Westinghouse study.) The scope of the evaluation is quite narrow. Many of the potentially beneficial aspects of Head Start were not even to be considered in the evaluation; other research groups were hired by OEO to evaluate different aspects of Head Start. These other studies included:

> (1) analyses of the immediate effects of full-year programs for 1966-67 and subsequent years; (2) an evaluation of Head Start's medical nutritional impact; (3) an assessment of the impact of Head Start programs on their communities; (4) a number of independent investigations of the short-term effects of selected Head Start projects in one location or area; (5) a recently mounted seven-year longitudinal study (p. 13).

In addition to the studies cited in the Westinghouse study, there has also been an evaluation of planned variation in Head Start programs.[1]

It is important to note that the Westinghouse study was to concentrate on the average impact of Head Start rather than on how effective the various approaches to Head Start were. Sheldon White (1970, p. 169) points out that OEO was sponsoring three types of evaluations in 1968. These are summative evaluations which "would assess the impact of a program as a whole, averaging over local variations in implementation"; formative evaluations which "would be concerned with locating and analyzing strong programs. . .to find ways to improve the overall effort," and monitoring which "would be. . .the periodic review of each local program to see to it that minimum expectations about performance were being met." From (3) in the list of goals we can see that the Westinghouse study was to be a summative evaluation. Although an evaluation of average impact might not be as interesting as the effects of specific types of programs to some researchers and policymakers, the information is important for comparing the effectiveness of the various antipoverty programs. White (1970, p. 13) points out that other summative evaluations were commissioned at about the same time to evaluate various manpower programs, Community Action Programs, Neighborhood Health Centers, and JOBS.

The RFP specified several other criteria for the evaluation. The study had to be done ex post facto rather than as a true experiment so that the evaluation could be completed within one year. The evaluation was to include several thousand children from approximately one hundred communities throughout the United States. In addition, the RFP specified that nonparticipating children were to be employed as a control group and that several instruments were to be used for measuring cognitive and affective development.

2. Plan of Research

The Westinghouse researchers felt that the primary question to be answered by the evaluation was:

> Does the cognitive and affective development of primary-grade school children who have had Head Start experience differ significantly from that of comparable children who have not had such experience? (p. 33)

There were other questions raised in the evaluation but these were considered secondary.

The Westinghouse researchers recognized the limitations of ex post facto studies but felt that an ex post facto study could be designed to adequately evaluate Head Start. Policy restrictions on Head Start enrollment precluded the use of a large scale true experiment; time and monetary constraints seemed to indicate that an ex post facto study was the correct procedure to use. The other studies previously cited were intended to gather information on Head Start that an ex post facto study could not obtain.

In Chapter 2 we demonstrated that nonrandom assignment can lead to bias in the measurement of treatment effect under certain conditions. The Westinghouse researchers attempted to eliminate bias by designating conditions that had to be met before a child could be considered a member of the Head Start or control population. These conditions are:

> (1) Continuity of residence in the target area. All subjects must have lived in the target area from the time of the specified Head Start program until the time of the study. . . .
>
> (2) Eligibility for Head Start. All subjects must have met the eligibility requirements for participating in the Head Start program.[2]
>
> (3) Equivalent school experience. All must have attended the same school system. . . .
>
> (4) No other Head Start experience. All those who had attended some other Head Start center or other similar preschool program were excluded (p. 36).

These four conditions established the Head Start and control subpopulations in each of the neighborhoods or "target areas" that were used in the study. Then a random sample of eight former Head Start participants was selected from each target area who were in each of the first, second, and third grades. Within each neighborhood a control sample of eight children at each grade level was then selected. The children in each control subpopulation were drawn randomly and were selected to match the Head Start sample on the basis of age, sex, kindergarten attendance, and racial/ethnic characteristics.[3] The groups were not matched on SES because the researchers felt that it would be too costly, but information on SES was collected and used as a covariate in the statistical analysis.

The Head Start neighborhoods to be used in the study were drawn randomly from a list of the 12,927 centers that existed in the 1966-67 school year. From the list of centers a sample of 100 and an oversample of 200 were drawn for the analysis. In both the primary sample and the oversample approximately 70 percent of the centers had summer programs and 30 percent had full-year programs; this represented the proportion of each that was found in the population of all centers. When a particular center in the primary sample could not be used in the study it was replaced with a center from the oversample which had the same type of program, i.e., summer or full-year. A total of 225 centers were screened before the final sample of 104 centers was established. The reasons for dropping the 121 centers are:[4]

> (1) Fifty centers were dropped because suitable control children could not be found within the target area.
>
> (2) Twenty-one centers were dropped because they had Head Start programs for one year only.
>
> (3) Nineteen schools declined to participate in the study for various administrative or scheduling reasons.
>
> (4) Eight schools lacked principals for summer participation.
>
> (5) Seven centers had Head Start populations too small for the study.

(6) Sixteen centers were dropped for miscellaneous other reasons (p. 40).

To determine if the omitted centers differed in any important ways from the final sample, an abridged form of the Head Start Official's Interview Questionnaire (HSIQ) was sent to the 121 dropped centers and the complete HSIQ was administered to the 104 centers in the final sample. Only 54 of the omitted centers replied, and their responses to the questions differed significantly at the 5 percent level on five of the 32 questions (pp. 42-43). There is no clear inference that can be drawn from these comparisons on whether or not we would expect the nonparticipating centers to be more effective; it would be important to know the characteristics of the nonresponding centers.

Because the Westinghouse study was designed to assess the average impact of Head Start throughout the United States, an attempt was made to determine if the children in the Westinghouse sample were comparable to the national population of Head Start children. Detailed information on all Head Start children was not available, but the U.S. Bureau of the Census surveyed approximately 25 percent of all full-year programs and 5 percent of all summer programs for 1966-67. It was discovered that the Westinghouse sample had a larger average family size (although the Westinghouse report does not say how much larger) and also a higher average family income. The only other major difference between the two samples was that Mexican Americans comprised a larger proportion of the Westinghouse sample.

3. Instruments Used in the Westinghouse Study

Three types of measurement instruments were used in the Westinghouse study: background data instruments, cognitive measures, and affective measures. The four background data instruments used in the study are the Head Start Official's Interview Questionnaire (HSIQ), the Parent Interview Questionnaire (PIQ), the Vocational Aspiration-Expectation Index (VAEI), and the School Environment Measure (SEM).[5] The HSIQ was used to gather information about the individual Head Start centers and the Head Start participants within each center. Topics covered by the HSIQ include the type of sponsoring agency, size and nature of the staff, racial composition of the participants, participation of parents in activities, number of children in the center, and the curriculum and objectives of the center. The HSIQ was administered in an interview conducted by a member of the Westinghouse staff with the Head Start official in charge of the center.

The PIQ is an 82-item questionnaire designed to obtain information about the home environment of the children in the sample. Included in the PIQ were questions on the home learning environment, parental attitudes toward the child, parental attitudes toward education, health of the child, vocational and educational aspirations for the child, and socioeconomic and demographic status of the family. Most of the information collected from the PIQ was not used in the primary data analysis, but the information for the SES covariate was formed by using information from the PIQ.

The VAEI was designed by the Westinghouse researchers to determine the educational and occupational aspirations and expectations parents had for their child. Although the VAEI is classified as a background instrument, it is conceivable that a successful Head Start program could alter parental aspirations and expectations. The VAEI has not been widely used nor is there an existing theoretical framework for including it in an evaluation of Head Start since the concepts it measures could be either causes or effects; therefore we shall not utilize it in our reanalysis. (The Westinghouse study does not report the VAEI scores for the control children, so it is impossible to determine if the groups differ on the variables measured by the VAEI.)

The final background instrument used in the Westinghouse study is the SEM. The SEM was devised by the Westinghouse researchers to measure "pertinent factors in the total school environment" (p. 66). The SEM concentrated on aspects such as the role of the principal, the emphasis on discipline, and the structural arrangement of the school. The SEM did not include questions on class size, teacher backgrounds, and curricula used by the schools which may be important factors in the educational production function. The SEM was not used as a control variable in the statistical analysis but was used solely to determine if the school environment varies greatly for children from different Head Start centers. The Westinghouse study concludes that "the variation in school environment in those schools serving Head Start children is not very great" (p. 126).

Two approaches were used in the study to measure cognitive development. Children were given an achievement test to determine how well they had mastered material appropriate to their grade level and an ability test to determine their scholastic aptitude and intelligence. As the level of achievement varies greatly within the primary grades, a different achievement test was used for each of the three grade levels. Children in the first grade were given the Metropolitan Readiness Test (MRT) to measure achievement. The MRT manual (Hilderth et al., p. 2) states that the MRT was "devised to measure the extent to which school beginners have developed in the several skills and abilities that contribute to readiness for first-grade instruction." The six subtests of the MRT are:

> Test 1. Word Meaning. A 16-item picture vocabulary test. The pupil selects from three pictures the one that illustrates the word the examiner names.
>
> Test 2. Listening. A 16-item test of ability to comprehend phrases and sentences instead of individual words. The pupil selects from three pictures the one which portrays a situation or event the examiner describes briefly.
>
> Test 3. Matching. A 14-item test of visual perception involving the recognition of similarities. The pupil marks one of three pictures which matches a given picture.
>
> Test 4. Alphabet. A 16-item test of ability to recognize lower-case letters of the alphabet. The pupil chooses a letter named from among four alternatives.
>
> Test 5. Numbers. A 26-item test of number knowledge.
>
> Test 6. Copying. A 14-item test which measures a combination of visual perception and motor control. (Hilderth et al., p. 2)

The MRT manual recommends that the total MRT score, the sum of the six subtest scores, be used for evaluating readiness. The manual cautions that "Efforts to attach significance to the subtest scores of individual pupils are not encouraged; the subtests are short, and so the reliabilities of their scores are naturally lower than that of the total score" (Hilderth et al., p. 19). The manual points out that the reliability coefficients of the subtests range from .50 to .86 and that the reliability coefficient for the total score is .91. Smith and Bissell (1970, p. 80) claim that the MRT is also a useful instrument because "the test is not only a measure of the child's performance upon entering first grade, but is often a determinant of the way he will be treated in the future" as the MRT is frequently used in reading group assignments. The MRT was administered to the children by their classroom teachers.

Achievement of children in the second and third grades was measured by the Stanford Achievement Test. Second grade children took the Primary I Battery (SAT2), and children in the third grade took the Primary II Battery (SAT3). The SAT manuals (Kelley et al., 1965, p. 2) state that the SAT is "a series of achievement tests developed to measure the important knowledges, skills, and understandings commonly accepted as desirable outcomes of the major branches of the elementary curriculum." The six subtests on the SAT2 are:

> 1. Word Reading. This test consists of 35 items, graduated in difficulty, which measure the ability of a pupil to analyze a word without the aid of context.
>
> 2. Paragraph Meaning. The Paragraph Meaning Test consists of a series of paragraphs, graduated in difficulty from which one or more words have been omitted. The pupil's task is to demonstrate his comprehension of the paragraph by selecting the proper word for each omission from four choices that are afforded him.
>
> 3. Vocabulary. The Vocabulary Test employs a multiple-choice type of item in which the pupil is required to select from a series of three

> alternatives the proper answer to a question or a statement read by the teacher.
>
> 4. Spelling. In this test the word to be spelled is pronounced by the teacher, an illustrative sentence is read, and the word is repeated, whereupon the pupil writes the word in his test booklet.
>
> 5. Word Study Skills. This test measures auditory perception of beginning and ending sounds, phonics, and phonograms.
>
> 6. Arithmetic. The Arithmetic Test measures the pupil's knowledge of measures, problem solving, and number concepts. (Kelly at al., 1966, pp. 4-5)

In contrast to the MRT, the SAT does not emphasize a total score, but the individual subtests have reliability coefficients that range from .79 to .95.

The SAT3 is similar in structure and content to the SAT2. The eight subtests of the SAT3 are:

> 1. Word Meaning. The Word Meaning Test consists of 36 multiple-choice items which measure the ability of the pupil to read a sentence silently to himself and then select the correct word to complete a sentence.
>
> 2. Paragraph Meaning. This test contains 60 items, each of which requires the pupil to select from among four printed words the one which fills an omission in the presented series of graduated paragraphs.
>
> 3. Science and Social Studies Concepts. This test presents 38 items of specialized vocabulary from these curricular fields and requires the pupil to demonstrate his knowledge of synonyms, definitions, and concepts.
>
> 4. Spelling. The Spelling Test consists of a 30-item dictation test.
>
> 5. Word Study Skills. This test measures the pupil's ability for auditory discrimination of beginning and ending sounds and his proficiency in visual phonics.
>
> 6. Language. The Language Test measures the pupil's knowledge of punctuation, capitalization, and grammar rules.
>
> 7. Arithmetic Computation. This test measures the pupil's skills in addition, subtraction, multiplication, and division.
>
> 8. Arithmetic Concepts. This test measures the pupil's problem-solving ability and understanding of number concepts. (Cicirelli et al., pp. 75-76)

The final measure of cognitive development used in the Westinghouse study is the revised edition of the Illinois Test of Psycholinguistic Abilities (ITPA). The object of the test, according to its authors, is to "delineate specific abilities and disabilities in children in order that remediation may be undertaken when needed" (Kirk et al., 1970, p. 5). The manual further states that "it is a diagnostic test of specific cognitive abilities, as well as a molar test of intelligence." The Westinghouse researchers regarded the ITPA quite highly: "From the inception of this project, the ITPA was regarded as valuable for obtaining a crucial kind of evidence for evaluating the effectiveness of Head Start" (p. 77). The ITPA has the advantage of being applicable for children from 2 to 10 years of age; the test was administered to children at all three grade levels and is the only cognitive instrument available that allows intergrade comparisons to be made. Not all educators share the Westinghouse researchers' enthusiasm for the ITPA.[6] The ITPA manual (Kirk et al., 1968) describes the ten subtests as follows:

> 1. Auditory Reception. This is a test to assess the ability of a child to derive meaning from verbally presented material.
>
> 2. Visual Reception. This test is comparable to the Auditory Reception Test but utilizes a different sense modality. It is a measure of the child's ability to gain meaning from visual symbols.
>
> 3. Auditory-Vocal Association. This test taps the child's ability to relate concepts presented orally. In this test the requirements of the auditory receptive process and the vocal expressive process are minimal while the organizing process of manipulating linguistic symbols in a meaningful way is tested by verbal analogies of increasing difficulty.
>
> 4. Visual-Motor Association. The organizing process in this channel is tapped by a picture association test with which to assess the child's ability to relate concepts presented visually.

5. Verbal Expression. The purpose of this test is to assess the ability of the child to express his own concepts vocally.

6. Manual Expression. This test taps the child's ability to express ideas manually. This ability is assessed by a gestural manipulation test.

7. Grammatic Closure. This test assesses the child's ability to make use of the redundancies of oral language in acquiring automatic habits for handling syntax and grammatic inflections. In this test the conceptual difficulty is low, but the task elicits the child's ability to respond automatically to often repeated verbal expressions of standard American speech.

8. Visual Closure. This test assesses the child's ability to identify a common object from an incomplete visual presentation.

9. Auditory Sequential Memory. This test assesses the child's ability to reproduce from memory sequences of digits increasing in length from two to eight digits.

10. Visual Sequential Memory. This test assesses the child's ability to reproduce sequences of nonmeaningful numbers from memory.

The ITPA was administered to the children individually by interviewers trained by the Westinghouse staff.

Although this dissertation shall not consider the affective benefits of Head Start, we shall briefly describe the three affective development instruments designed and utilized by the Westinghouse study. First is the Children's Self-Concept Index (CSCI), a 26-item test designed to "assess the degree of positive self-concept of children in the primary grades" (p. 81). The Classroom Behavior Inventory (CBI) is the second affective measure devised for the study. Its main function is "to assess the children's motivation to achieve in school learning" (p. 83). The third affective instrument developed for the Westinghouse study is the Children's Attitudinal Range Indicator (CARI) which is designed "to assess the child's positive and negative attitudes toward peers, home, school, and society" (p. 85). The CSCI and CARI were administered by classroom teachers, and the CBI was completed by each child's regular teacher. It is difficult to assess the reliability and validity of the affective instruments as they were standardized on groups of 100-200 students.

4. Statistical Model

In Chapter 2 we presented several statistical models that could be employed for Head Start evaluation. The Westinghouse researchers used three statistical models, but they emphasized one particular analysis of covariance model while referring to the others as "alternate models." We will present their basic model in the equivalent regression framework because economists are more familiar with regression techniques, and because it is easier to discern the underlying structure of the model in the regression framework. The basic equation of the model is

$$Y = \beta_0 + \beta_Z Z + \beta_1 X_1 + \sum_{j=2}^{N} \beta_j X_j + u$$

where the variables are defined as:

Y = the score on the cognitive or affective tests

Z = a dummy variable for experimental status where
$Z = 1$ for the children who participated in Head Start, and
$Z = 0$ for the children in the control group

X_1 = the measure of SES used in the analysis

$X_2 \ldots X_N$ = a series of dummy variables for the N neighborhoods or target areas where
$X_j = 1$ if the observation is from neighborhood j, and
$X_j = 0$ otherwise

u = a disturbance term assumed to be normally distributed with a mean of zero, and a constant variance, and assumed to be uncorrelated with the independent variables.

The Westinghouse researchers assumed that the data should be grouped by neighborhood rather than use individual children as the unit of observation; for each neighborhood there were thus two observations--one for the Head Start children and one for the control children. The mean value for each group on each variable is determined and used for the value of that observation. Because there are N neighborhoods, each with an experimental and control group, there is a total of 2N observations for the analysis. The Westinghouse report does not explain why grouped data rather than ungrouped data were used in the model. The model also assumes that the regression is linear in nature with Head Start, SES, and neighborhood as the appropriate independent variables and that there is no interaction effect between Head Start and SES or neighborhood.

The variable used for SES in the Westinghouse study is the "Two-Factor Index of Social Position" described in Hollingshead (1958). The variable is constructed as a weighted sum of the head of household's occupation and educational attainment. Occupation is computed on a seven-point scale as follows:

1. Executives and proprietors of large concerns, e.g., doctor, lawyer, commissioned officer, etc.

2. Managers and proprietors of medium-sized businesses and lesser professionals, e.g., police chief, registered nurse, teacher.

3. Administrative personnel of large concerns, owners of small independent businesses, semiprofessionals, e.g., clothing shop owner, IBM programmer, florist, etc.

4. Owners of little businesses, clerical and sales workers, and technicians.

5. Skilled workers, e.g., barber, baker, fireman, policeman, hair stylist, painter, construction foreman, etc.

6. Semi-skilled workers, e.g., truck driver, nurse's aide, practical nurse, hairdresser, housekeeper, enlisted military, etc.

7. Unskilled workers, e.g., laundry worker, farm hand, garbage collector, construction worker, unemployed, welfare (p. B-29).

The educational component of the Hollingshead Index is also formed by using a seven-point scale:

1. Graduate school
2. Completed college
3. Some college
4. High school graduate
5. Some high school
6. Seventh to ninth grade
7. Less than seventh grade (p. B-28).

The Hollingshead Index score is computed by multiplying the occupational score by seven, multiplying the educational score by four, and taking the sum of the two products. The Hollingshead Index is a very crude measure of SES, and one can question its use as a control variable for SES.[7]

The inclusion of the neighborhood variables was justified in the report as follows: "It was felt that the use of a model such as this, in which variation due to replications could be removed from the general error term, would provide a more sensitive test, since the variation from target area to target area was large" (p. 103).

The dependent variable, Y, is used to represent each of the dependent variables available for a particular sample. Thus for the first grade children there were 21 regressions to be run as there are 21 dependent variables available: the three affective tests (CBI, CSCI, and CARI), the six subtests plus the total score of the MRT, and the ten subtests plus the total score for the ITPA.

Within each equation the estimate of the effects of Head Start upon the dependent variable is given by the partial regression coefficient β_Z. The coefficient is then tested for statistical significance by using the t- statistic associated with the coefficient.

5. Results of Data Analyses

There is a great deal of data analysis that was carried out by the Westinghouse researchers, and we shall only attempt to summarize the findings here. Each of the cognitive and affective test scores was used as a dependent variable for full-year and summer samples for the first, second, and third grades. In addition, the samples were stratified further by geographic region (Western, Northeastern, and Southeastern), racial/ethnic characteristics of the centers (mainly black, mainly white, mainly Mexican American, and miscellaneous) and by type of population units (large cities, suburbs, small cities near large cities, other small cities, and rural). In most cases, the null hypothesis that Head Start has no effect on the cognitive development of children cannot be rejected by the analyses performed by the Westinghouse researchers. For summer Head Start programs the report states:

> With regard to language development, the following observations can be made:
> 1. For the overall analysis, there were no significant differences between Head Start and control, either on the ITPA "Total Raw Score" or on the subscores at any of the three grades.
>
> 2. For the analysis made by subgroups, there were no differences in favor of Head Start on "Total Raw Score" in any of the groupings, and only a few scattered subscore differences (in the Southeast, mainly Negro, and Mexican-American groupings).

3. For the analysis of subgroups, there were differences in favor of the controls on "Total Raw Score" (as well as some subscores) in the Northeast and the mainly white centers at grade 1, and in the West and in the centers in small cities far from core cities at grade 3. There were some scattered subscore differences as well.

With regard to school readiness and achievement, the following observations are made:
1. With regard to the MRT (in grade 1), significant differences were found on the overall analysis in favor of controls for two subscores. On the subgroup analyses, controls were superior to Head Start in the Northeast ("Total Readiness" and two subscores), as well as five scattered subscores in other groupings.

2. With regard to the SAT at grade 2, there were significant differences found in favor of controls on the "Battery Median" and some subscores, both on the overall analysis and in the Western area, as well as for three scattered subscores.

3. With regard to the SAT at grade 3, there were no differences found in the overall analysis, but there were significant differences in favor of Head Start on the "Battery Median" and on one subscore in the subgroup analysis for core cities and for mainly Negro centers (p. 140).

The analyses for the full-year Head Start programs tended to indicate that full-year Head Start is somewhat effective:

1. ITPA, grade 1. No significant differences existed between Head Start and control in "Total Raw Score," when the group of centers was analyzed as a whole or by subgroups. Some significant differences in subscores existed in certain subgroups of centers.

2. ITPA, grade 2. No significant differences were found in "Total Raw Score" for the overall analysis, although significant differences were found for "Visual Sequence" and "Manual Expression." For the Western and Southeastern groupings of centers, there were significant differences in "Total Raw Score" and various subscores (especially "Visual Reception" and "Auditory Association"). When centers were grouped according to city size or racial/ethnic composition, a few significant differences in subscores were found. [The significant differences reported in this section are for a positive treatment effect.]

3. ITPA, grade 3. No significant differences were found for the particular group of six centers tested. Because these same six centers showed no difference at grade 1 or grade 2, no "wash out" or "catch up" trend could be inferred.

4. MRT, grade 1. In the overall analysis, there was a significant difference in favor of Head Start on "Total Readiness" and on the "Listening" subscore. In addition, there was a statistically significant difference in favor of former Head Start enrollees from centers in core cities in readiness, not only in total score, but in most of the subscores.

5. SAT, grade 2. For the overall analysis, there was no significant difference between Head Start and control groups, either for the "Battery Median" or for subscores. When centers were grouped according to geographic location, there was a significant difference om "Battery Median," "Word Reading," and "Arithmetic" [in favor of Head Start] for the Southeastern centers. Differences [also in favor of Head Start] in "Arithmetic" were also found in two other subgroupings.

6. SAT, grade 3. No significant differences were found for the particular group of six centers tested. These same six centers showed no differences at grade 1 or grade 2 (p. 157).

Several points should be kept in mind when interpreting these findings. First, the "differences" reported in the summaries above are not the differences in the raw scores but are the differences adjusted for the SES and neighborhood covariates; in the regression framework these differences are the coefficients for treatment. Second, the number of observations that these analyses employ is small, especially for the full-year cases. For the overall summer analysis there were 98 observations for grade 3, 136 for grade 2, and 142 for grade 1. When the summer sample is stratified by racial/ethnic composition, population unit, or geographic location the number of observations is reduced to as few as 16 observations. In the full-year analyses there are 54, 56, and 12 observations for the first, second, and third grades, respectively. In the stratified full-year analyses the number of observations ranges from 12 to 22; because the total full-year third grade sample was so small, no stratified analyses were carried out.

(The reason for the small number of observations for all the analyses is that the data were grouped.) For the summer samples, approximately half of the coefficients which were not significant were positive; most of the insignificant coefficients were close to zero. For the full-year samples, most of the Head Start coefficients that were not statistically significant indicated a positive Head Start effect, but they also tended to be small.

The Westinghouse researchers realized the great difficulties involved in interpreting their results: "There is, then, no final, definitive and indisputable answer to the basic question of when an educational intervention program can be deemed effective, successful, and worthwhile, and when it can not. We do not propose or pretend to offer one here" (p. 163). The Westinghouse report does offer a method of determining if the effects of Head Start are of "practical significance":

> The measure of practical significance can be summarized as follows. Using the standard deviation of the population on which a given test is standardized as a reference, if the difference between means due to an instructional treatment (e.g., the difference on any selected measuring instrument between Head Start and control group) is at least one-half of the aforementioned standard deviation (0.5 SD) or more, then the magnitude of such a difference may be considered of sufficient practical relevance and worthwhileness (p. 164).

This criterion is arbitrary, and in Chapter 4 we shall consider alternative measures of practical significance. The Westinghouse report defends this criterion by citing three references which recommend its use.[8] Using this criterion of practical significance the Westinghouse report concludes that:

> In summary, when one looks at the observed effects of Head Start according to the test of practical relevance, it must

> be concluded that the effects found on the standarized tests are indeed small in magnitude, with the exception of a few differences found in subgroups of full-year centers on the ITPA, and do not meet the criterion of practical relevance (p. 168).

Several alternate techniques for analyzing the data were also employed by the Westinghouse researchers in case the assumptions of the covariance model were not correct. The first alternate analysis utilized the Wilcoxon matched-pairs signed-ranks test (also referred to as the Wilcoxon one-sample test). In this procedure a treatment effect is estimated for each Head Start center and the Wilcoxon test is used to determine if the distribution of the treatment effects is symmetric about zero.[9] The advantage of the Wilcoxon test is that no assumption is made about the normality of the residuals. The main disadvantage of the procedure is that it produces no estimate of the magnitude of the effect of Head Start. The Wilcoxon analysis produced results similar to those of the covariance model--Head Start appeared to be marginally effective for full-year centers but ineffective for the summer centers.

The second alternate analysis used is analysis of covariance for individuals rather than for groups. This method was tried because the Westinghouse researchers felt the increase in degrees of freedom could increase the power of the statistical tests to detect small but significant effects of Head Start. The three covariates that were included in this analysis (in addition to a variable for Head Start) are per capita income of the child's family, occupational status of the head of the household, and educational attainment of the head of the household. The covariates were included as precoded variables rather than as sets of dummy variables; thus, the occupational status variable was coded as 7 if the head was unskilled, 6 if he was semiskilled, etc., using the scales from the PIQ described on page 59. Per capita income was divided into the following seven categories: 1. \$200-399, 2. \$400-599, 3. \$600-799, 4. \$800-999, 5. \$1000-1499, 6. \$1500-2499, 7. \$2500-5000. The Westinghouse researchers claim that one weakness of this procedure is the omission of the neighborhood variables from the analysis, but there is nothing to preclude their use as the number of observations has increased eightfold. The analysis of covariance for individuals produced results that were similar to the analysis of covariance for groups; summer Head Start again appeared to be ineffective and full-year Head Start was marginally effective.

6. Conclusions of the Westinghouse Study

Virtually all of the evidence presented in the Westinghouse report indicates that summer Head Start programs have been ineffective and full-year programs have been marginally effective in raising the level of cognitive development for preschool children. It is therefore not surprising that the report concludes with the following specific recommendations:

> 1. Summer programs should be phased out as early as feasible, and converted into full-year programs or extended year.
> 2. Present full-year programs should be continued but every effort should be made to render them more effective (pp. 247-248).

The reader should keep in mind that these conclusions are based on the analysis and interpretations of the Westinghouse researchers;

we have reservations about the statistical procedures used and about the interpretations given in the report. In this chapter we have sought only to summarize the Westinghouse report. We shall offer criticisms and describe our reanalysis in the next chapter.

FOOTNOTES

[1] The Westinghouse study did not provide references for these other studies, and I have been able to locate only a few of them. Many of the small-scale evaluations are reviewed by Datta (1969) and Grotberg (1970). An interim report for the planned variation study has been written by Bissell (1971). The study of the impact of Head Start on communities was released in May 1970 and can possibly be obtained under the following reference: Kirschner Associates, Inc. A National Survey of the Impacts of Head Start Centers on Community Institutions, a report presented to the Office of Child Development, U.S. Department of Health, Education and Welfare pursuant to contract B89-4638, May 1970.

[2] Although OEO has eligibility criteria for Head Start participation, these criteria are very broad and individual Head Start centers could use discretion to make the requirements more stringent. The Westinghouse researchers learned the requirements within each target area and restricted their control population to those who met these requirements.

[3] The text of the Westinghouse report is contradictory about the characteristics on which the control and Head Start children were matched. On page 37 the report states: "the control sample was matched to the Head Start sample on the following variables known to affect school performance: sex, racial/ethnic group membership, and whether or not kindergarten was attended." However, on page 127 the report states that "the Head Start and control children were matched for age, sex, and kindergarten attendance." In our reanalysis of the data we have found that the groups were matched on all four demographic characteristics rather than either set of three characteristics.

[4] Smith and Bissell (1970) have strongly criticized the screening procedure used. Their criticisms will be considered in Chapter 4.

[5] Copies of the background data instruments are in Appendix B of the Westinghouse report.

[6] Smith and Bissell (1970, p. 80) suggest that the ITPA's "validity when used with disadvantaged preschool children has not yet been established and its reliability with these children is questionable." They cite published articles which have indicated that the ITPA may be very unreliable for disadvantaged preschool children.

[7] We shall comment in detail on the Hollingshead Index in Chapter 4. At this point we call attention to the arbitrary weighting factor, the omission of income in a variable purporting to measure economic status, and the arbitrary and confusing ranking of the occupational scale (i.e., construction workers have a lower score than household domestics, and "owners of small independent businesses" are given a different score than "owners of little businesses." Note that the Hollingshead Index is an inverse measure of SES where a high score is indicative of a low SES.

[8] The three references cited by the Westinghouse study on the 0.5 SD criterion are: Jacob Cohen, "Some Statistical Issues in Psychological Research," in *Handbook of Clinical Psychology*, B. Wolman ed. (New York: McGraw-Hill, 1965), p. 101; A. A. Lumsdaine, "Instruments and Media of Instruction," in *Handbook of Research in Teaching*, N. L. Gage ed. (Chicago: Rand McNally, 1963), p. 609; and G. Manly, *Psychology for Effective Teaching* (New York: Holt, Rinehart & Winston, 1960), pp. 309-401.

[9] The procedure for using the Wilcoxon test is as follows: first, the children in a given target area are stratified by some SES variable; second, the mean of the control group on the dependent variable for each category is subtracted from the mean of the Head Start group in the same category; third, a weighted average of the differences, d, is computed for the target area where the weights are the proportion of children from the target area within each category. This procedure is repeated for the N target areas, producing N d-statistics. The d's are then rank-ordered in terms of their absolute size with the rank of 1 given to the smallest d and N to the largest d. The ranks are then given the sign of d, and the test statistic "T" is defined as the smaller of the sums of the like-signed ranks. For large samples

$$\frac{T - N(N + 1)/4}{[N(N + 1)(2N + 1)/24]} - 1/2$$

is distributed normally. [See Brownlee (1965, p. 258).]

Chapter 4

Criticisms of the Westinghouse Study

1. Introduction

The Westinghouse study became controversial even before it was officially released. On April 17, 1969, an article in the New York Times appeared: "Head Start Report Assailed as Full of Holes and Potential Political Disaster." In this article Dr. William G. Madow, a statistician who served as a consultant for the Westinghouse study, is quoted as saying, "Far from accepting payment for my work, I would appreciate your removing my name as a statistical consultant, since I would not want anyone to think I had any responsibility for the design--and the analyses and conclusions based on it seem to me to be incorrect." In the same article Dr. Alfred Yankarer of the Harvard School of Public Health claims that the Westinghouse report has "many, many scientific holes," and that it would be "a perversion of science" to use the report as a guide for policy making. The report was defended by Dr. Victor G. Cicirelli, its principal author, and by several officials at OEO. On April 27, 1969, Daniel Patrick Moynihan, President Nixon's urban affairs advisor at that time, stated in an interview with the New York Times that "The Report is a good one. It tells us that a popular program which we thought was working may not be working so well after all." Within a year after the Westinghouse report's publication, critical articles by Campbell and Erlebacher (1970), Smith and Bissell (1970), and Madow (1969) were published. Each of these articles was responded to by Cicirelli, sometimes with the collaboration of John W. Evans and Jeffrey Schiller of OEO. Articles defending the Westinghouse study were also written by Walter Williams and Evans (1970) of OEO, and by Sheldon H. White (1970), who served as a consultant for the study.

Virtually every aspect of the Westinghouse study has been criticized, and the criticisms can be divided into five broad categories:

1. Were the questions to be answered by the Westinghouse study the correct ones?

2. Were the cognitive, affective, and background instruments utilized in the study appropriate and reliable for an evaluation of Head Start?

3. Were the sampling procedures used in the Westinghouse study for the selection of Head Start centers and individuals correct to locate a representative sample, or would they necessarily produce a biased sample?

4. Were the statistical procedures employed in the Westinghouse study correct or would they lead to biased estimates of effects of Head Start?

5. Were the interpretations and conclusions of the study correct or were they inappropriate and possibly misleading?

We shall examine each of these questions to assess the validity of the findings of the Westinghouse study.

2. The Focus of the Westinghouse Study

Criticisms of the central questions that the Westinghouse study attempted to answer can be divided into two categories: (1) the study completely ignored all potential benefits of Head Start other than cognitive and affective benefits; and (2) the Westinghouse study did little research concerning whether or not some Head Start centers were effective and what attributes of Head Start would lead to success. Madow (1969, p. 251) phrases the problem as follows: "Did the objectives of the Westinghouse study correspond to the nature and importance of the program being evaluated? My answer is negative."

The specific objectives of the Westinghouse study were not determined by the Westinghouse researchers but were listed in OEO's Request for Proposals (RFP); the RFP stated that the study was to assess "Head Start impact, on the average, across the country . . . " (p. 14). The Westinghouse researchers recognized the importance of measuring all of the benefits to determine if Head Start is worthwhile. One of the first points made in the report is that the Westinghouse study was one of many evaluations of Head Start being undertaken at that time, and that other evaluations were designed to assess the medical and nutritional aspects of Head Start, the impact of Head Start on communities, and the long-term benefits of Head Start by using a seven-year longitudinal study. In addition, there is a planned variation project currently being carried out that will attempt to determine the most effective approach to preschool education (see Bissell (1971)).

What remains to be answered is whether or not the questions asked in the Westinghouse study are fruitful and if policy implications can be drawn from such a broad evaluation. Although the medical and nutritional aspects of Head Start are important components of the program, these benefits could easily be provided at lower cost by a custodial day care center or an expanded public health service; as Sar A. Levitan states (1969, p. 133), "Head Start was sold to the American public as an expanded kindergarten program for the poor, with health and nutrition components." In his reply to Madow, Evans (1969, p. 254) states:

> . . . while Head Start has objectives other than cognitive and affective change, these other objectives are in large part instrumental to the cognitive and affective objectives. That is, the program is attempting to improve children medically and nutritionally in order to make it possible to change them cognitively and motivationally.

For a benefit-cost analysis we should include the medical and nutritional benefits of Head Start, but since the goal of improving cognitive development is what differentiates Head Start from day care programs we would be especially interested in discovering the cognitive benefits of Head Start.

The determination of what leads to a successful Head Start program is an important one, especially to educators and researchers. Some critics feel that by asking whether or not the average Head Start program produces cognitive benefits we may ignore another important question of "can Head Start produce cognitive benefits?" In her review of preschool programs, Marion S. Stearns (1971, p. 125) phrases the problem this way:

> That is, if there were a few Head Start programs making tremendous differences in children's later achievements in the primary grades, these would not be recognized by a study such as that by Westinghouse. Although they would count toward raising the average overall performance scores presumably, their weight in a large group would be small.

The above statement is correct, and any social scientists or policymaker interested in learning how education can be efficiently produced would be interested in a planned variation study. However, in a program that has been deliberately decentralized to allow for a variety of approaches to preschool education, the matter of the average impact of the program is important. In the previous chapter we mentioned that the Westinghouse study was to be a summative evaluation whose function was to "assess the impact of the project as a whole," rather than a formative evaluation which "would be concerned with locating and analyzing strong programs . . . to find ways to improve the overall effort." White (1970, pp. 178-179) gives an excellent description of the justification for summative evaluations: government officials must somehow decide which antipoverty programs are producing results that justify their costs so that government funds can be efficiently spent.[1] The fears of some critics that the Westinghouse report's negative conclusions would lead to the abandonment of preschool education programs by the Nixon Administration were unfounded; indeed, an April 27, 1969, New York Times story states:

> Mr. Nixon has said nothing disparaging about Head Start and indeed appears to have decided to have another go at making it work. But because of the Westinghouse report, he has said he will treat it as "experimental" rather than as an established program of final answer.

Thus we conclude that the questions addressed by the Westinghouse researchers are valuable ones and can be used for making policy decisions. We also note that these are not the only questions about Head Start that should be considered, but they were questions that could be answered in a short time period to get preliminary information about the average impact of Head Start.

3. The Instruments Used in the Westinghouse Study

There has been some criticism of the cognitive and background instruments used in the Westinghouse study and a great deal of criticism of the affective instruments that were developed for and employed in the study. Although our reanalysis of the Westinghouse data is defined by the instruments originally employed, we shall point out the weaknesses and flaws of these instruments and mention how they weaken and limit the analyses that can be carried out.

The affective measures developed for the Westinghouse study, the *Children's Attitudinal Range Indicator* (CARI), the *Children's Self-Concept Index* (CSCI), and the *Classroom Behavior Inventory* (CBI), have been criticized because the instruments were validated on vary small samples (from 100 to 200 children) and because there is little evidence that the tests accurately measure the psychological attitudes in question. The Westinghouse researchers and others associated with the study do not defend the affective instruments very strongly. In their reply to the Smith and Bissell article, Cicirelli, Evans, and Schiller (1970, p. 115) state that "our judgment about the affective findings should be tentative and this is the view the Westinghouse Report took." In Evans's reply to Madow (1970, p. 256), he says, "No great claims are made for the affective instruments." White, who served as a consultant to the Westinghouse study, generally defended the research and conclusions of the study in his review article (1970); he dismisses *all* of the affective findings and says that "the affective instruments were not good. . . ."[2] Because of the weaknesses of the affective instruments and our lack of expertise in interpreting the scores on these instruments, we shall confine our reanalysis of the Westinghouse data to the cognitive benefits of Head Start. In any event, we believe that cognitive development is a more important goal of Head Start than affective development.

The cognitive instruments used in the Westinghouse study (the ITPA, MRT, and SAT) have not been subjected to as much criticism. All of these instruments were developed and widely used prior to the Westinghouse study; these same instruments have been commonly used in other evaluations of preschool and primary grade education programs. Smith and Bissell (1970, p. 80) argue that the ITPA may not be an appropriate instrument for evaluating the cognitive benefits of Head Start because "its validity when used with disadvantaged preschool children has not yet been established and its reliability with these children is questionable." Cicirelli, Evans, and Schiller (1970, p. 115) respond that the children in the Westinghouse sample were not of preschool age at the time of the testing and that high reliabilities have been found for both normal and retarded children on the ITPA. They further argue that:

> Moreover, considering the facts that the ITPA deals in a multifactoral way with the crucially important variables of language and that it is individually administered by trained examiners, it would seem to us that an argument could be made that ITPA is at least as good a cognitive measure as the MRT for the purposes of this study, perhaps better.

Four background instruments were developed for the Westinghouse study, but the information collected from two of the instruments was not used in the analysis of the effects of Head Start. The Vocational Aspiration-Expectation Index (VAEI) was apparently not considered important, as there is no comparison of the VAEI scores for the Head Start and control children; if this instrument is robust, it could conceivably have been used to determine if the two groups had similar expectations for their children, and we could determine if the groups differ on parental attitudes. We are not sure whether the VAEI should be classified as a dependent or an independent variable so we shall not utilize it in our reanalysis. The School Environment Measure (SEM) could have proved to be a useful instrument for children in the second and third grades if specific questions about the curriculum, class size, and level of training of teachers had been included; these variables would have made the Westinghouse study useful as a study of school inputs in the educational production function. Unfortunately, the questions on the SEM were vague and cannot be used in this manner; therefore the SEM will not play any part in our reanalysis of the Westinghouse data.

The Head Start Official's Interview Questionnaire (HSIQ) contained information about the racial and ethnic composition of the centers, and the Westinghouse study used this information to stratify the centers to determine if centers with certain racial characteristics are more effective than others. The HSIQ also contains some data that can be used to compare different Head Start programs. Information is available about the child-teacher ratio, general approach of the center, sponsoring agency, etc., but the HSIQ does not contain certain crucial information such as the number of hours per day children spend at the center and the budget of the center. Nevertheless we feel that to exploit the Westinghouse data to the fullest possible extent, a study of the effects of Head Start center variation should be attempted.

The Parent Interview Questionnaire (PIQ) was administered to one of the parents of each child in the sample to obtain information on the home learning environment, parental attitudes toward the child, parental attitudes toward education, the health of the child, parental vocational and educational aspirations for the child, and socioeconomic and demographic information about the child's family. Of this information only the educational and occupational information was used to construct the Hollingshead Index of socioeconomic status. This variable was used as a covariate in the statistical analysis. Some of the shortcomings of the PIQ are:

1. When the PIQ was completed by someone other than the child's mother or father, it is impossible to determine the number of parents present and the educational and occupational levels of the individual parents. Thus, about seven percent of the observations had to be skipped due to unavailable data.

2. The information on total family income is only available in broad categories; e.g., less than $2000, $2000-3999, etc. It would have been more useful if income were given exactly and broken down by source such as father's wages, mother's wages, public assistance payments, and other sources of income.

3. The occupational codes are too broad and ambiguous. For example, "owners of small independent businesses" are ranked higher than "owners of little businesses." The rankings are not always consistent with socioeconomic status; for example, housekeepers are ranked higher than construction workers.

4. The information on employment status is applicable only for the time when the PIQ was administered. There is no information about previous work history.

4. Sampling Procedures

Two types of criticisms of the sampling procedures utilized by the Westinghouse study have been made. The first criticism is that the centers selected for the study are not representative of all Head Start centers, and thus, the Westinghouse study did not assess the average impact of all Head Start centers. The second criticism is that the selection of the individual Head Start and control children was done in such a way that the two groups were unequal in ability (especially cognitive ability) prior to their Head Start experience. This second criticism implies that the statistical analysis will lead to biased estimates of the effects of Head Start under some conditions.

Smith and Bissell argue that the sampling process used for the selection of centers was inappropriate for several reasons. They claim that stratified random sampling would have been preferable for determining the relative effectiveness of different kinds of centers. This is correct, but the Westinghouse study only attempted to differentiate between full-year and summer programs. Thus, simple random sampling is the appropriate method to determine the *average* impact of Head Start. Because the full-year and summer samples were never combined, there was no reason for selecting fewer full-year centers. The number of full-year centers was too small to permit an extensive analysis of subgroups. Replacing summer centers with full-year centers also reduces the efficiency of the estimate of the effects of summer Head Start while increasing the efficiency for full-year Head Start; if we are equally interested in the two types of programs, an equal number of each type of centers would be the most desirable assuming that the populations served by the two types of programs are similar. Evans (1969, p. 255) recognizes this point: "On retrospect this was an erroneous and unnecessary decision since we decided relatively early we would at no time combine the summer and full-year samples. If we were doing the study over, we would select a larger number of full-year centers."

Smith and Bissell also emphasize the fact that 225 centers needed to be screened before the final sample of 104 was selected. They cite the reasons for the omission of centers and claim that the final sample is biased to overrepresent less effective centers. They claim that the 21 centers dropped because they only had programs in 1967 creates bias because "the sample was most representative of those centers which were funded in the early days of Head Start, and any idiosyncracies in the allocation of the early funds were carried over into the study," (p. 67). Later in the same article, they claim that "Head Start centers are likely to be both of longer duration and relatively more effective during their second year of operation . . . than during their first year of operation" (p. 98). They obviously cannot criticize the study from both angles. The second argument appears to have more a priori appeal.

A group of 27 centers was dropped because arrangements could not be made for the children to take the tests in the school or because the school was closed during the summer making it impossible to obtain the chldren's records. It is doubtful that these conditions would introduce any systematic bias into the sample of Head Start centers. Three of the reasons for dropping centers could possibly have led to the elimination of especially effective centers. It is conceivable that the 50 centers that were dropped because there were too few eligible control children were very effective and, hence, were able to recruit all eligible children. The seven centers that were dropped because they produced too few Head Start graduates may have been better because they were small (as is argued by Smith and Bissell), but because they represent only a very small fraction of the total Head Start population it is unlikely that a large bias would be created by their omission. Sixteen centers were dropped for miscellaneous reasons such as "the center admitted only retarded children, there was dissension between the school and the center, etc." (Cicirelli, Evans, and Schiller, p. 110). Dissension between the center and school could be a sign of either an effective or ineffective center, but there is no evidence available for either view.

Although there is no a priori reason to expect that the dropped centers were more effective than the included ones, the Westinghouse researchers anticipated criticism for dropping so many centers and sought to demonstrate that there were few differences between the two groups by mailing an abridged version of the HSIQ to the dropped centers. Because the HSIQ was administered orally to the directors of the participating centers it is possible that some response bias was produced by the different form of administration. It is difficult to make meaningful comparisons between the participating and dropped centers on the basis of the HSIQ because only 54 of the 121 dropped centers returned the questionnaire. It could be argued that those centers that did return the HSIQ were most interested in helping to evaluate Head Start and thus more effective than those that did not respond, but this is a weak argument that cannot be supported by any evidence. When the responses of the two groups were compared, the Westinghouse researchers found significant differences on five of the 32 items, and these differences indicate that the participating centers are more likely to be effective. For example, the study found that a higher proportion of the participating centers followed up their graduates' performance in first grade. Smith and Bissell make the point that the procedure used to determine if the two groups differed was crude, but the fact remains that there is no indication that the final sample of 104 centers overrepresented ineffective centers.

The selection of the individual children for the Head Start and control groups has been criticized on the grounds that the control children came from a more advantaged background; in Chapter 2 we demonstrated that under some conditions a difference of this nature will lead to underestimates of the effects of Head Start in the statistical analysis. This is the major criticism of Campbell and Erlebacher (1970), and Madow (1969). Although the Head Start and control children were matched on the basis of neighborhood, Head Start eligibility, age, race, sex, and kindergarten attendance, there was no attempt to match the samples on SES. Analysis of the control and Head Start samples confirmed that the control children were slightly more advantaged than the Head Start children on several SES measures. We shall discuss the implications of these differences below. To the extent that the Head Start children were self-selected, a case can be made for bias in favor of the Head Start group; parents who are the most interested in helping their children would tend to be parents who are more interested in their cognitive development.

5. Criticisms of the Statistical Analysis

The statistical procedures utilized in the Westinghouse study have been criticized mainly on the basis that the analysis of covariance technique used in the analysis does not adequately control for initial differences in cognitive development in the Head Start and control groups. As the Westinghouse data were collected ex post facto, there is no measure of cognitive ability prior to the Head Start experience. The SES information that was collected can be used as a proxy for prior ability, but this measure is fallible. In Chapter 2 it was demonstrated that when certain selection procedures are used for assigning children to the Head Start and control groups the use of a fallible measure of pretreatment cognitive development will lead to a biased estimate of the effects of Head Start; under these conditions if the Head Start population is less able than the control group (i.e., a "scraping" selection procedure was used) then the analysis of covariance will lead to an underestimate of the effect of Head Start. The Campbell-Erlebacher model is one model that will lead to an underestimate of the effect of Head Start, but several of the alternative models do not lead to bias. Because the children in the control group appear to come from slightly higher SES backgrounds, Madow, and Campbell and Erlebacher argue that the Westinghouse study's findings are biased to underestimate the effect of Head Start.

In Chapter 2 we noted how discriminant analysis can be employed to determine if the Head Start and control groups differ significantly on their SES. We shall use discriminant analysis in our reanalysis of the Westinghouse data, but we must realize its limitations for determining if the regression analysis will lead to biased estimates of the effect of Head Start. These limitations were discussed in Chapter 2, so they need not be repeated in detail. The most important limitation of discriminant analysis is that there are several models where the groups will differ significantly on their SES but where regression analysis with SES variables included as regressors will not lead to biased estimates of the coefficient of Head Start.

The model presented in Chapter 2, where group selection is determined by pretest score, is one case where the Head Start and control children differ on their pretreatment levels of cognitive development but where regression analysis will produce an unbiased coefficient for Head Start. This model is not exactly relevant for our reanalysis of the Westinghouse data because the Head Start selection was not generally made in this manner, and if it were, we do not have information on what screening device was used.[3] The omitted variable model presented in Chapter 2 can also be used to demonstrate that observed differences in SES may not lead to a biased estimate of treatment effect. As we have previously noted, parental attitudes toward education are unmeasured, and we would expect these attitudes to have an effect independent of SES on the cognitive development of the child. Since we would expect those parents who enrolled their children in Head Start to be stronger in these attitudes, the omission of parental attitudes will introduce a bias toward overstating the effects of Head Start. It is impossible to predict the magnitude of this bias.[4] In our discussion of omitted-variable models in Chapter 2 we also noted that the child's pretreatment level of cognitive development could be considered an omitted variable. We then demonstrated that if selection within SES levels is random with regard to ability then a regression of posttest score on SES and a treatment variable will introduce no bias in the estimate of treatment effect, even if lower SES levels are more highly represented in the Head Start group.

In retrospect, it is apparent that a major weakness of the Westinghouse study is that no firm statement can be made about the bias problem. It would have been helpful if the HSIQ contained detailed questions concerning the selection procedures used at the Head Start centers. Because different recruiting procedures may have been used at different types of centers the bias may be of different magnitudes even of a different direction for various subsamples. For example, it might be that the Westinghouse findings that the predominantly black full-year centers were somewhat effective was because the center officials used a "creaming" selection process whereby the center admitted the best of the eligible children. It is impossible to rectify this problem for the Westinghouse data, but we recommend that future ex post facto evaluations make a strong effort to learn the selection procedures used in the program so that any bias can be taken into account when interpreting the results.

Although the issue of bias has been the major criticism of the Westinghouse study's statistical analysis, other questions have also been raised. Smith and Bissell make some interesting charges against the Westinghouse study that deserve a more complete response than they received from Cicirelli, Evans, and Schiller. They begin by dismissing the argument that the groups are much different on the basis of SES:

> The means and standard deviations for the four socioeconomic variables [parents' income, parents' education, parents' occupation, and number of children] show only small differences between the Head Start and control groups. It is probably unreasonable to anticipate that these differences would be of crucial importance, particularly since their effect should be controlled in the data analysis through a covariance procedure (pp. 74-75).

(It should be noted that Smith and Bissell are referring only to the full-year first grade sample and that the differences between the Head Start and control groups were greater for the summer samples.) They then argue that the correlation coefficients between the SES variables and the MRT total scores are significantly different for the two groups and that this will lead to bias. This is puzzling. The covariance model employed by Smith and Bissell, essentially a regression of the MRT score on the four SES variables and a dummy variable for Head Start, does not require that the correlation matrix be the same in the two groups. Smith and Bissell assert that the differences in the correlations imply that there is an interaction between Head Start and SES upon the MRT score as well as an additive effect. The possibility of such an interaction effect is worth considering, but there is no need to justify such a hypothesis from the correlation coefficients. Smith and Bissell tested the hypothesis that there is an interaction effect and found that the hypothesis was supported. They do not report the procedure used for testing the hypothesis, and it is likely, considering the results of our reanalysis of the same data, that they applied the wrong test. The procedure outlined at the conclusion of the article by Gregory Chow (which they cite) provides a test for determining if the constant as well as the regression coefficients are equal in two samples. Since a nonzero coefficient for Head Start would lead to a different constant for the two regressions, it is inappropriate to include the constant in the test of equality.

If there is an interaction effect between Head Start and SES we can no longer consider the value of Head Start in terms of a single number; the effect will depend on the child's initial level of SES. This is illustrated in the diagram below:

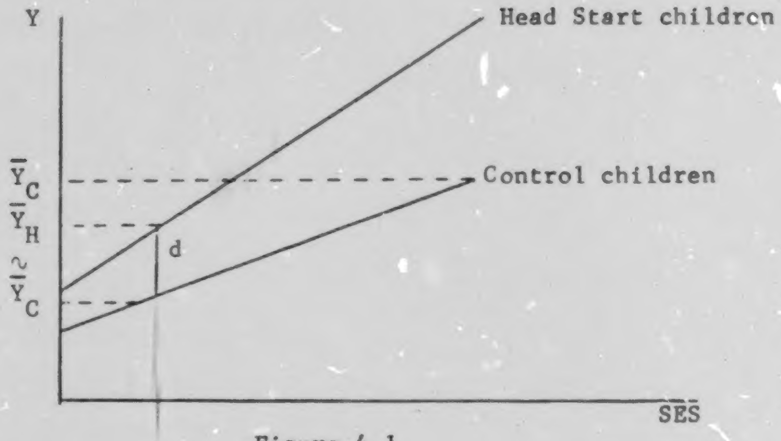

Figure 4.1

Head Start-SES Interaction

For simplicity, we assume that SES can be represented by a single variable. An interaction effect is illustrated in the diagram because the regression lines for the Head Start and control groups are not parallel. The effect of Head Start for a child is the vertical distance between the two regression lines at his level of SES. In Figure 4.1, $\bar{Y}_H$ represents the mean of the MRT for the Head Start children and $\bar{Y}_C$ the MRT mean for the control children. Smith and Bissell define the effect of Head Start, d, to be the vertical distance between the lines at the point where $\bar{Y}_H$ is on the Head Start regression line. The value $\tilde{\bar{Y}}_C$ is the MRT score we would predict the average Head Start child would have received if he had not participated in Head Start. This procedure is incorrect because the effect of Head Start depends in part on the SES of the child, and it is misleading to summarize the effect as one value.

The primary analyses in the Westinghouse study used only the Hollingshead Index, the dummy variables for neighborhood, and a dummy variable for Head Start as independent variables; in addition the data were grouped by neighborhood. We believe that several improvements can be made in the structure of the statistical model employed. Whatever the merits of the Hollingshead Index are in measuring status, there is no theoretical reason for including socioeconomic information in that form for a model of cognitive development. The variables that are used in computing the Hollingshead Index, the occupational and educational achievement of the head of the household, can be included in the regression as separate sets of dummy variables. This procedure offers several advantages: (1) the Hollingshead Index constrains the regression coefficient for education achievement to be four-sevenths of the coefficient for occupational status, whereas including the variables separately permits the coefficients to vary freely;[5] and (2) the precoding of these variables constrains the model by requiring that the effect upon the cognitive development of a child is the same when a parent's score is increased from any given step to the one above; thus, the model implies that an increase of the head's education from 0-6 years of education to 7-9 would have the same effect on the cognitive development of a child as a shift from 9-11 years to 12 years, ceteris paribus. As our knowledge of educational production functions is very limited, such arbitrary constraints add nothing to the model and may reduce its usefulness in our attempts to discover the effects of SES upon cognitive development.

Variables not included in the Hollingshead Index are also appropriate for inclusion in the regression model. For two-parent families the education and occupation of the head's spouse is appropriate and should add to the explanatory power of the model as well as partly control for differences in ability between the Head Start and control groups--thereby reducing the potential bias problem. In their reanalysis of the Westinghouse data, the procedure used by Smith and Bissell (1970, p. 74) of using only the higher of the two parents' educational levels is arbitrary, and we believe that the backgrounds of both parents will be important. Although family income is given only within seven broad brackets, we believe that some measure of income must be included as an independent variable; in the Westinghouse study's alternative analysis, a single precoded variable was used, but other conceivable approaches are to use the midpoints of the brackets or a per capita income measure.

If the model is analyzed with individuals as the observation units we can include demographic variables such as age, race, sex, and kindergarten attendance in the model. In Chapter 5 we shall define all of the variables utilized in our reanalysis and our hypothesis about each of them. At this time we only wish to note that many relevant variables were excluded from the Westinghouse analysis and that many of those that were included were included in an arbitrary form. Improvements in the model where we better measure the socioeconomic background of the children will reduce bias caused by differences in the Head Start and control backgrounds; in effect, adding relevant SES variables gives us a more accurate "pretest."[6]

Our final suggestion for the statistical analysis is that individual data rather than grouped data should be used. The Westinghouse report gives only a weak justification for using grouped data, and there are several reasons to prefer using ungrouped data. Cramer (1964, p. 237) has demonstrated that in a two-variable model, grouping by the independent variable produces unbiased estimates of the regression coefficients but the estimates are not as efficient as when ungrouped data are used. Blalock (1961, pp. 102-112) reaches the same conclusions and also notes that if the grouping is done on the dependent variable our estimates of the regression become biased. Blalock also considers the case where the grouping is by geographic proximity, which is the case for the Westinghouse study. In the empirical case he has selected to illustrate the problem, grouping by proximity is similar to grouping by the independent variable; but he warns:

> In other instances, however, a grouping by proximity might approach more closely a grouping by the dependent variable. In this latter case, we might be badly misled by the numerical value of b_{yx} which would then increase with the size of the grouping (p. 112).

A final reason for using ungrouped data is that many variables of interest cannot be included in the model when grouped data are used. For example, we may be interested in testing the hypothesis that child's sex has no effect on the test scores. When we use ungrouped data we can test this hypothesis by including a dummy variable for sex; but, when grouped data are used there is no variation across neighborhoods in the proportion of males and females so we must omit this variable from the analysis. If we wish to use sets of dummy variables for characteristics such as educational attainment, there is no way to do so when grouped data are employed. Thus, we shall carry out all of our reanalyses using ungrouped data.

6. Interpretations of Statistical Findings and Strategies for Evaluations

We now consider the difficult task of making interpretations of the statistical analyses of social action programs such as Head Start. To be useful to policymakers an evaluation must include specific recommendations based on the analysis of the data. For an evaluation of Head Start this includes: (1) the determination of our best estimate of the effects of summer and full-year Head Start on various groups of children; (2) the consideration of how confident we can be in the accuracy of our estimates; (3) the determination of whether the benefits of the program exceed the costs; and finally, (4) a comparison of Head Start to policy alternatives to determine if Head Start is the most efficient program for meeting our goals. If the first and second steps indicate a zero or negative effect of Head Start, the subsequent steps are unnecessary, and Head Start can be viewed as an ineffective program.

The first step in interpreting our results is to examine the regression coefficient for the effect of Head Start; in the Westinghouse study where there are many cognitive tests available we have effects of Head Start on each of these measures, but we shall make the simplifying assumption that there is only one dependent variable. If the effects of Head Start appear to be zero or negative (and we are reasonably confident that our estimate is unbiased), our work is done--we may conclude that in its present form Head Start is an inappropriate program for increasing the cognitive development of disadvantaged children. If the estimate is positive we then consider the statistical significance of our estimate. If the hypothesis of no effect cannot be rejected at the 1 or 5 percent significance level it is sometimes argued that the program should be abandoned because there is no "significant" effect. This argument is rarely correct. As Cain and Watts (1970, p. 233) explain, "a body of data may be unable to reject the hypothesis that some coefficient is zero and be equally consistent with a hypothesis embodying a miraculously high effect." We must keep in mind that our estimate is the best estimate of treatment effect and that a large standard error for the coefficient only implies that we cannot be very confident in the accuracy of the estimate. If we find that the effect of Head Start is not statistically significant and the effect of the program appears large enough to be considered worth attaining, a more refined evaluation of Head Start should be attempted. Some of the techniques that can be used to increase the efficiency of the estimates are to increase sample size, improve the accuracy of measurement of the independent variables, and select the sample so that there is less covariation between Head Start and the independent variables.

A Head Start instructor may be satisfied to learn only what the cognitive benefits of Head Start are, but a policymaker must know more. The policymaker must know whether the benefits of Head Start exceed the costs of the program so that he can determine if it is worth the investment. This involves placing a dollar value on cognitive gains--a very difficult task, but one which must be faced. If it costs $1,000 to make Head Start available to one child, then the policymaker must decide if the gains produced by Head Start are worth at least $1,000 for the program to be considered worthwhile. Economic theory asserts that an individual will purchase private market goods only if he decides that it is worth at least the cost to him. For publicly funded social action programs, however, it becomes much more difficult to price the benefits, especially when the benefits are in the form of gains in cognitive development for young children. Presumably the benefits of Head Start will also include day care benefits to the parents and externality benefits to the taxpaying public, but for our illustration we assume that we are only interested in the value of the cognitive benefits.

After we have expressed the benefits of Head Start in dollar terms we compare the benefits to the costs of the program. If the costs exceed the benefits, then Head Start should not be continued in its present form even if it produces significant benefits. For example, we would reject a job training program that increased the present value of lifetime earnings by $10,000 if the program cost $15,000. It is important to include all benefits in the calculations, however, because even if the primary educational benefits do not exceed the costs, it is possible that the value added from the externality benefits and the day care benefits may bring the total dollar benefits to a sum greater than the total costs.

The final step in interpreting our results is to compare the efficiency of Head Start with alternative policies. To illustrate this point we shall use the following hypothetical example. We assume that our analysis includes three policy variables: Head Start (Z), kindergarten (K), and cash transfers (I). We further assume that kindergarten and Head Start are continuous variables and that the level of treatment can vary continuously. Children can receive any combination of the treatments, but we assume that the relationship between the test score (Y) and all independent variables is linear and additive with no interaction effects.

We begin the analysis by regressing Y on the three policy variables and all other independent variables that our model specifices. Suppose that our fitted regression equation is

$$Y = \beta_0 + .0005I + 1.00Z + 5.00K + \sum_{i=1}^{N} \beta_i X_i$$

where the X_i variables represent the nonpolicy variables such as race, age, and number of siblings. Because the regression coefficient for Head Start (Z) is positive, we can proceed to compare the costs and benefits of the program. For this example we assume that one unit of Head Start costs $1,000 per child If the policymaker decides that a gain of one point for a child is worth $1,000 then the program can be judged a success. The policymaker can then compare Head Start with the alternative programs to see which is the most efficient method of increasing the cognitive development of disadvantaged children. Suppose that we have the following cost information:

cost of a unit of Head Start = \$1,000

cost of a unit of kindergarten = \$2,000

cost of income transfer = \$1.

By combining the cost information with the regression coefficients we can arrive at the benefit-cost ratios for the three policies:

points/dollar for Head Start = .0010

points/dollar for kindergarten = .0025

points/dollar for income transfer = .0005.

Thus for the example we have presented, Head Start is not the most efficient way of increasing the cognitive development of children; kindergarten provides more benefits per dollar of expenditure.[7]

The procedure for determining the appropriate strategy becomes more complex if we introduce political constraints or if we allow Head Start and kindergarten to be provided at only one treatment level. We might have found that transfer payments would be more efficient, but transfer payments may not be so politically acceptable as an educational program. (In economic terms we could say that transfer payments create negative externalities and educational programs create positive externalities and that our benefit figures should be adjusted accordingly.) If we do not permit the treatment level for kindergarten and Head Start to vary, we have an interesting equity-efficiency problem. When the budget for educational programs is fixed, it is no longer a simple matter to determine if the money should be spent on Head Start or on kindergarten. In our example we would produce the largest aggregate amount of benefits by concentrating all of the funds on kindergarten; yet this procedure will provide benefits to a smaller number of children and may be criticized on equity grounds. The policymaker must then decide how to weight the equity and efficiency aspects in making his decision.[8]

The sole criterion advocated in the Westinghouse study for determining the "practical significance" of Head Start is that if the gain in cognitive development attributed to Head Start is at least as great as one-half of one standard deviation on the tests used in the study, then Head Start can be considered to have significant benefits. This measure suffers because it is not only arbitrary (why not one-quarter of one standard deviation as Smith and Bissell suggest, or one full standard deviation?) but it also ignores many of the important issues that should be considered.

7. Summary

In this chapter we have tried to enumerate the shortcomings of the Westinghouse study and to provide suggestions for future evaluations. Some of the problems of the Westinghouse study can be corrected by a reanalysis of the data; other problems are impossible to rectify at this time. In Chapter 5 we present our reanalysis of the data and note the problems that remain. It is unfortunate that some important issues, including whether or not regression analysis will lead to unbiased estimates of treatment effects, cannot be settled. As the number of large-scale evaluations such as the Westinghouse study continue to grow, we hope that more attention is paid to the area of evaluation methodology.

FOOTNOTES

[1] White (1970, pp. 178-179) defends summative evaluation as follows:

> Why be concerned with so gross and uninformative a question in the first place? There is a question I once heard an OEO official ask:
>
> 'We are paying 350 million dollars for this program and 20 million dollars for that program. Why are we paying 350 million dollars for this program and 20 million dollars for that program? Why aren't we paying 20 million dollars for this program and 350 million dollars for that program?'
>
> This is the question of the policymaker. It is a new kind of question, deceptively like the traditional forms of the evaluation question, but different and worthwhile in its own terms. Is national implementation of Head Start not working?--Does it produce effects substantial enough to justify its cost and its preemption of funds which might be allocated toward other possible approaches?

[2] Although Cicirelli, Evans, and Schiller (1970) feel that the affective measures can be used for making "tentative" conclusions, White (1970, p. 173) argues that they should not have been used at all:

> Though the affective measures were ultimately used to come to conclusions in the main body of the draft report, I am inclined to feel that this is a little misleading. I believe that the contractor could have pleaded nolo contendre about the issue of attitude measurement.

[3] Because selection for Head Start was decentralized with each center doing its own selection, it is very unlikely that all centers used the same criteria.

[4] Some evaluations of preschool education programs have eliminated this bias by selecting the control group as well as the experimental group from volunteers for the program. Even in this case it is possible that neither group is representative of all eligible children, and the results cannot be extrapolated for other children. Note that if the program administrator "scrapes" on the basis of variables negatively correlated with parental attitudes the bias due to the omission of these variables would be offset to some extent.

[5] The coefficient for educational level is constrained to be four-sevenths of the coefficient for occupation because the Hollingshead Index is formed as a weighted average of the two components, with the weights being 4 and 7, respectively.

[6] Consider, for example, the model presented in Chapter 2, section 4. From equation (25) it can be shown that the bias in β_Z will be smaller as the accuracy of the pretest increases.

[7] Cain and Watts (1970, p. 238) make the same point about determining the most efficient policy alternative. In their presentation, however, they suggest scaling the policy variables so that one unit of any policy variable has the same cost. If this procedure is used then the most efficient policy variable is the one with the largest regression coefficient. In our example the regression equation would be:

$$Y = \beta_0 + .0005I + .0010Z + .0025K + \sum_{i=1}^{N} \beta_i X_i$$

and once again kindergarten would be the most efficient program.

[8] A discussion of this general topic is included in the article by Weisbrod (1970).

Chapter 5

Empirical Findings

1. Introduction

In this chapter we present the empirical results of our reanalysis of the Westinghouse data. Because there are many results to report, only the most important findings are included in this chapter, and an appendix is included which reports and discusses some of our other results.

As we have stated earlier, Head Start can be viewed as one input factor in an educational production function. It is possible that the parameters for an educational production function may vary for different groups of individuals, so it is necessary to stratify the sample and estimate a separate function for each group. We have stratified the data by grade level (first, second, and third), type of Head Start program (full-year and summer), and by number of parents present (both and mother only). There are ten rather than twelve subsamples to be analyzed because there were insufficient observations for the third grade full-year categories. Stratification by grade is necessary because we wish to examine the effects of Head Start separately for each grade level and because different cognitive tests were administered to children at each grade level. We have stratified by type of Head Start program because the control children were selected to match the Head Start children separately for each of the two types of Head Start, and combining the two samples might introduce bias into the analysis if the selectivity is different for summer and full-year programs. Stratification by parents present was done because the set of independent variables is different when both parents are present rather than when only the mother is present.

One of our criticisms of the Westinghouse study is that no attempt was made to incorporate most of the socioeconomic and demographic data collected into the analysis. We have expanded the vector of independent variables to include the available variables that might influence performance on the measures of cognitive development. Many of these variables are included as sets of dummy variables to allow for a more general functional form. By expanding the set of independent variables we expect to reduce the possible bias in the coefficient for Head Start caused by nonrandom treatment assignment. In addition, the variables we have included are useful in helping us to learn more about the educational production process, and several of the variables can be used as policy variables in addition to or in place of Head Start.

A list of the variables employed in our reanalysis and a description of how the variables were formed is given in Table 5.1. For each subsample analyzed two regressions were run. The first regression employed the mean for ITPA scores as the dependent variable and the second regression employed either the mean for MRT scores, SAT2 scores, or SAT3 scores depending upon the grade level. Regarding the independent variables, when a single trait, such as mother's education, is represented by a group of dummy variables (i.e., MSOCOL, MHSG, MSOHS, M79, and M06) one of the variables must be omitted from the regression so

Table 5.1

Description of Variables Used in the Reanalysis

Variable	Description
CHILD	Number of children in the family
INCOME	Total annual income of the child's family in dollars
AGE	Age of the child to the nearest year
MSOCOL	1 if child's mother has more than 12 years of education; 0 otherwise
MHSG	1 of child's mother has 12 years of education; 0 otherwise
MSOHS	1 if child's mother has 10-11 years of education; 0 otherwise
M79	1 if child's mother has 7-9 years of education; 0 otherwise
M06	1 if child's mother has 0-6 years of education; 0 otherwise
MOPRO	1 if child's mother has professional or managerial occupation; 0 otherwise
MOCLER	1 if child's mother has clerical occupation; 0 otherwise
MOSKIL	1 if child's mother has skilled occupation; 0 otherwise
MOSEMI	1 if child's mother has semiskilled occupation; 0 otherwise
MOUNSK	1 if child's mother has unskilled occupation or no occupation; 0 otherwise
FEMALE	1 if child is female; 0 otherwise
MALE	1 if child is male; 0 otherwise
RURAL	1 if child lives in a rural area; 0 otherwise
KIND	1 if child attended kindergarten; 0 otherwise

Table 5.1 (cont.)

Variable	Description
NOKIND	1 if child did not attend kindergarten; 0 otherwise
FSOCOL	1 if child's father has more than 12 years of education; 0 otherwise
FHSG	1 if child's father has 12 years of education; 0 otherwise
FSOHS	1 if child's father has 10-11 years of education; 0 otherwise
F79	1 if child's father has 7-9 years of education; 0 otherwise
F06	1 if child's father has 0-6 years of education; 0 otherwise
FAPRO	1 if child's father has professional or managerial occupation; 0 otherwise
FACLER	1 if child's father has clerical occupation; 0 otherwise
FASKIL	1 if child's father has skilled occupation; 0 otherwise
FASEMI	1 if child's father has semiskilled occupation; 0 otherwise
FAUNSK	1 if child's father has unskilled occupation; 0 otherwise
WHITE	1 if child is white; 0 otherwise
BLACK	1 if child is black; 0 otherwise
MEXAM	1 if child is Mexican American; 0 otherwise
HDSTRT	1 if child has had Head Start; 0 otherwise
BLKHS	1 if child is black and has had Head Start; 0 otherwise
ITPAMN	mean of child's nonzero scores on ITPA
MRTMN	mean of child's nonzero scores on MRT

Table 5.1 (cont.)

Variable	Description
SAT2MN	mean of child's nonzero scores on SAT2
SAT3MN	mean of child's nonzero scores on SAT3
DIVOR	1 if child's parents divorced; 0 otherwise
SEPAR	1 if child's parents separated; 0 otherwise
WIDOW	1 if child's mother is a widow; 0 otherwise
NEVMAR	1 if child's mother never married; 0 otherwise

that the normal equations for the regression coefficients can be solved. Where there are more than two categories we have generally followed the procedure of omitting the variable representing the highest category; for mother's education we have therefore omitted the variable for mothers who have had one or more years of college (MSOCOL) from the regressions. The category for which no dummy variable is included is called the reference group, and the coefficient for another variable in the set tells us how much higher or lower we would expect the child to score on the dependent variable if he is in that category rather than in the reference group, other things equal. For the subsamples where both parents are present we have omitted the set of variables for marital status since these categories do not apply; for similar reasons we have omitted the variables for father's occupation and education for the subsamples where only the mother is present.

There were some problems with the Westinghouse data, and we shall explain how we resolved them. Some of the observations lacked information on one or more of the independent variables, and these observations have been omitted from our reanalysis. For the dependent variables (ITPAMN, MRTMN, SAT2MN, and SAT3MN) we have used the mean of the nonzero subtests because we felt that a zero score indicates that the child had not really taken the subtest. We omitted all observations where the child had a zero score on three or more subtests of a given test because either the child had effectively not taken the test or he was so severely retarded that he did not belong in the study. The data for total family income was coded in brackets and we have used the midpoints of the brackets to assign values to the variable INCOME.

2. Hypotheses

We shall now describe our hypotheses concerning the coefficients of the independent variables. Our hypotheses deal with the sign of the coefficients and sometimes with relative magnitude, but we do not have enough knowledge about the educational production process to make stronger hypotheses. We expected the coefficient for the number of children in the family (CHILD) to be negative for two reasons. A greater number of children in a family with income held constant implies that the family's per capita income is lower; thus as family size increases the family will have less resources for providing a better learning environment in the home. Jensen (1969, pp. 72-74) claims that the intelligence of a child tends to be lower as the number of older siblings is increased; if this is true we have an additional reason for expecting the coefficient for number of children to be negative. We hypothesize that annual family income (INCOME) will have a positive coefficient because income not only reflects the capability of the parents to provide a good home environment, but also may be an indicator of the parents' cognitive ability which will presumably affect the child's ability through inheritance. There are some problems with the measure of income available from the Westinghouse data which might tend to lead to lower coefficients for income. The relevant measure of income for our study is permanent income rather than reported income. It is possible that the variables included in the model for occupational and educational status of the parents provide a better measure of permanent income than reported family income. If the parents included welfare transfer payments in their reported income the available measure will be subject to bias, especially because most of the families in our sample have a low level of income; this implies that the data on observed income overstates the appropriate value and will thus bias the coefficient toward zero. There are reasons why the age of a child (AGE) could have either a positive or negative coefficient, but we expect the arguments for a positive coefficient to dominate. Many of the communication skills measured by the ITPA (such as auditory reception and visual reception) are not fully developed in five- or six-year old children; we expect this would lead to a positive coefficient for age when the mean of ITPA scores is the dependent variable. Age can also be expected to be positively associated with the more achievement-oriented MRT because older children have had more time to increase their vocabulary and learn to count, which are some of the skills tested by the MRT. Working in the opposite direction, however, is the fact that some of the older children may have been prohibited from entering first grade a year earlier due to their low ability. We suspect that few of the children were held back from entering school at the appropriate time, and therefore that the coefficient for age will be positive.

The education of a child's parents should be positively associated with his cognitive development for several reasons. To the extent that the education attainment of the parents reflects their intelligence, the inheritance process will lead to a positive effect for education of parents on their children's cognitive development. The parents' level of education may also serve as a proxy for the parents' interest in their children's education, and more highly educated parents may therefore offer their children more encouragement for learning. Although these expectations are appropriate for the population as a whole, the

Westinghouse data is not representative of all families. Only children who were eligible for Head Start were considered for the study, and because Head Start programs sought to enroll disadvantaged children the parents with a college education must be atypical of the general population of college-educated parents; these parents may be considered as "failures" because their children are classified as disadvantaged in spite of the college education of the parents. The implication of this sample truncation is that children whose parents are in the base groups for the educational attainment variables may not only be expected to score lower on the cognitive tests than children of other college-educated parents but also lower than children whose parents are in some of the lower categories. In our regressions the variables for parental education are included as two sets of dummy variables (one for the mother and one for the father) with the highest categories (one or more years of college) serving as the reference group. We therefore expect the coefficients within each set to decrease in value as we proceed from higher to lower levels of education. If there were no truncation problem we would expect all of the coefficients in these groups to be negative, but we now expect that some of the higher categories might have positive coefficients.

Our expectations for the coefficients of the parents' occupational status variables are similar to those for the parents' education variables--we expect children whose parents are in more skilled occupations to do better on the cognitive tests because their parents are likely to be more skilled and thereby provide more positive reinforcement for education in the home. There may be a truncation problem similar to that discussed for the education variables, and there is also a strong possibility that some of the observations in the highest categories (professional and managerial workers) were miscoded or otherwise, the children would not be eligible for Head Start. For these reasons we may again find that children whose parents are in the reference groups may score lower on the cognitive tests than children whose parents are in some of the lower categories. We therefore expect the coefficients within each set to decline in value as we move from higher to lower levels of occupational status, and because of the sample truncation and possible miscoding, the coefficients for the higher categories in each group (clerical and skilled workers) may be positive. There is one other unusual feature of the information on parents' occupation. After examining the data it was realized that some of the mothers who were not in the labor force were reported as having an unskilled occupation and others reported that they had no occupation. This inconsistency was resolved by combining the two groups and considering them all as unskilled. Thus, if a child's mother is classified as unskilled it might indicate that the mother is not in the labor force and will have more time at home to spend with her child. This could lead to finding the coefficient for unskilled mother (MOUNSK) to be greater than the coefficients for some of the other categories.

We expected the coefficient for the dummy variable for females (FEMALE) to be zero (indicating that male and female children from similar backgrounds will not score differently on the tests) but we included this variable in case there would be differential development or if one or more of the tests discriminate against one sex. There is no way to determine if a significant coefficient for the sex dummy variable is indicative of differences in cognitive development or bias in the structure of the test. The variable for rural area residence (RURAL) was included to determine if children living in rural areas perform differently than comparable nonrural children; we had no a priori expectations about the coefficient of this variable.

Black and Mexican American children have been observed to score lower than white children on many ability and achievement tests. We lack the expertise to determine if this difference is due to real differences in ability or to cultural bias in the tests. It should be noted that in most of our subsamples the black and Mexican American children come from significantly more disadvantaged backgrounds than the white children.[2,3] Moreover, it is likely that discrimination against these minority groups can lower their performance, and no quantitative adjustment is available for this factor. We therefore expect coefficients for the ethnicity variables (BLACK and MEXAM) to be negative.

Two of the most interesting variables in the regressions are kindergarten attendance (KIND) and Head Start (HDSTRT). Kindergarten is available as an additional policy variable that can be used as a complement or substitute for Head Start. We expect kindergarten to have a positive effect on cognitive development, and we are especially interested in comparing the coefficients for kindergarten and Head Start. If kindergarten and Head Start both have positive effects upon cognitive development, one could examine the relative cost-effectiveness of the two policies.

It is conceivable that Head Start could be more effective for some ethnic groups than others, and several alternatives are available for examining this possibility. One approach is to stratify the sample by ethnic group and to run separate regressions for each group. A second approach is to combine the ethnic groups and to add dummy variables, such as one for blacks who have had Head Start, which is set equal to one only for children who are in the ethnic group and who have had Head Start. If we believed that Head Start must have the same effect for all ethnic groups we could pool the ethnic groups and include only one variable for Head Start. The first approach allows all of the coefficients in the educational production function to vary across groups. If the second approach is used all of the variables except for the constant and for Head Start are constrained to have the same effect for all ethnic groups; for the third approach all variables (except for the constant) are constrained to have the same effect for all ethnic groups. In our preliminary work we have tried all three approaches, and it appears that the second approach is best. When the samples were stratified by race, the coefficients for the socioeconomic and demographic variables did not differ substantially across ethnic groups, so it is more efficient to pool all of the ethnic groups for estimating these parameters; the only variable that consistently had a different coefficient is kindergarten, so we have included a dummy variable for blacks who have attended kindergarten (BLKIND). To anticipate our results, we have also found that there was often a great difference in the effects of Head Start for blacks and Mexican Americans compared to whites, so we have added dummy variables to permit the effect of Head Start to differ by ethnic group.

3. Empirical Findings for the Primary Samples

We shall now examine the regression equations that we have fitted for some of the more important samples to determine how well our hypotheses have been supported. We shall examine the results of the first grade subsamples where both parents are present in detail because we are most interested in discovering the effect of Head Start for these subsamples; a discussion of the regressions for the other subsamples is contained in the appendix to this thesis. The means and standard deviations for these samples can be found in Table 5.2 and Table 5.3. The information is given separately for the Head Start and control children within each sample so that comparisons between the groups can be made. It is readily apparent that the children in these samples are from disadvantaged backgrounds. The average number of children is greater than four and average family income is under $6,000 for both samples. The relevance of the differences between the Head Start and control children within each sample will be discussed later in this chapter. The regression equations where the mean of ITPA scores is the dependent variable are contained in Table 5.4, and the regressions where the mean of MRT scores is the dependent variable are in Table 5.5. The coefficients of determination (R^2) for our models range from .19 to .44 indicating that a maximum of 44 percent of the variation in the dependent variable can be explained by the independent variables. The coefficients of determination reported in this chapter are in the same range as those found in similar studies; for example, Samuel Bowles (1970) reports coefficients of determination that vary from .09 to .31 in his estimates of an educational production function for twelfth grade students, even when variables measuring school environment and student attitudes are included.

Table 5.2

Means and Standard Deviations for Grade 1, Summer, Both Parents Present Sample

Variable	Head Start		Control	
	Mean	Standard Deviation	Mean	Standard Deviation
CHILD	4.69	2.10	4.12	2.04
INCOME	5049.	2514.	5859.	2986.
AGE	5.89	.446	5.95	.548
MHSG	.263	.441	.416	.494
MSOHS	.362	.481	.237	.426
M79	.194	.396	.192	.394
M06	.092	.290	.065	.247
MOCLER	.109	.137	.042	.201
MOSKIL	.032	.176	.023	.149
MOSEMI	.130	.337	.078	.268
MOUNSK	.794	.405	.834	.372
FEMALE	.511	.501	.490	.501
RURAL	.263	.441	.234	.424
FHSG	.225	.419	.295	.457
FSOHS	.222	.416	.286	.452
F79	.263	.441	.175	.381
F06	.165	.372	.123	.329
FACLER	.041	.199	.045	.209
FASKIL	.197	.398	.240	.428
FASEMI	.311	.464	.351	.478
FAUNSK	.403	.491	.263	.441

Table 5.2 (cont.)

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
BLACK	.289	.454	.247	.432
MEXAM	.133	.340	.107	.310
KIND	.625	.485	.614	.488
BLKIND	.194	.396	.175	.381
ITPAMN	19.13	3.52	19.30	3.74
MRTMN	9.03	2.60	9.31	2.74
N	315		308	

Table 5.3

Means and Standard Deviations for Grade 1, Full-Year, Both Parents Present Sample

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
CHILD	4.83	2.17	4.74	2.01
INCOME	4861.	2252.	5490.	2656.
AGE	5.97	.492	5.96	.420
MHSG	.385	.489	.304	.462
MSOHS	.308	.464	.294	.458
M79	.192	.396	.196	.399
M06	.067	.252	.088	.285
MOCLER	.019	.138	.039	.195
MOSKIL	.019	.138	.029	.170
MOSEMI	.173	.380	.157	.365
MOUNSK	.769	.423	.735	.443
FEMALE	.510	.502	.539	.501
RURAL	.212	.410	.245	.432
FHSG	.308	.464	.304	.462
FSOHS	.260	.441	.196	.399
F79	.173	.380	.147	.356
F06	.192	.396	.157	.365
FACLER	0.00	0.00	.029	.170
FASKIL	.240	.429	.284	.453
FASEMI	.356	.481	.363	.483
FAUNSK	.337	.475	.245	.432

Table 5.3 (cont.)

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
BLACK	.500	.502	.480	.502
MEXAM	.192	.396	.157	.365
KIND	.558	.499	.588	.495
BLKIND	.288	.455	.304	.462
ITPAMN	18.84	3.84	19.66	4.26
MRTMN	8.74	2.53	8.81	2.68
N		104		102

Table 5.4

Effects of Individual Characteristics for Grade 1, Both Parents Present Sample, on Child's ITPA Score, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	-.134* (-1.900)	-.355** (-2.720)
INCOME	-.00001 (-.17000)	.0004** (3.2900)
AGE	.766** (2.790)	1.44** (2.52)
MHSG	-.194 (-.350)	-.291 (-.260)
MSOHS	-.508 (-.870)	-.552 (-.480)
M79	-.736 (-1.150)	-1.22 (-.97)
M06	-1.62** (-1.97)	-.512 (-.330)
MOCLER	2.42** (1.97)	-2.95 (-1.39)
MOSKIL	1.88 (1.51)	-.901 (-.400)
MOSEMI	.158 (.150)	-1.13 (-.70)
MOUNSK	.506 (.520)	-1.80 (-1.16)
FEMALE	-.120 (-.430)	.384 (.750)
RURAL	.240 (.700)	1.17 (1.03)
FHSG	.341 (.690)	-1.73* (-1.90)
FSOHS	-.003 (-.010)	-.950 (-.990)

Table 5.4 (cont.)

Independent Variable	Summer	Full Year
F79	-.475	-2.16*
	(-.890)	(-1.99)
F06	-.905	-1.07
	(-1.450)	(-.95)
FACLER	-.921	1.25
	(-1.080)	(.52)
FASKIL	-1.01	.576
	(-1.63)	(.490)
FASEMI	-1.17*	-.221
	(-1.90)	(-.190)
FAUNSK	-1.64**	-.040
	(-2.57)	(-.030)
BLACK	-2.10**	.810
	(-3.11)	(.560)
MEXAM	-1.12*	-.469
	(-1.69)	(-.330)
KIND	.862**	3.47**
	(2.500)	(3.94)
BLKIND	-.593	-3.56**
	(-.900)	(-3.11)
HDSTRT	-.361	-1.22
	(-1.000)	(-1.38)
BLCKHS	1.99**	2.04*
	(3.13)	(1.79)
MEXHS	1.53*	.081
	(1.78)	(.050)
CONSTANT	16.67**	12.00**
	(1.98)	(2.82)
R^2	.186	.374
N	623	206

Note: t-statistics are in parentheses below their coefficients.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

Table 5.5

Effects of Individual Characteristics for Grade 1, Both Parents Present Sample, on Child's MRT Score, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	-.073 (-1.430)	-.199** (-2.530)
INCOME	.00003 (.71000)	.0002** (2.7900)
AGE	.674** (3.380)	.732** (2.120)
MHSG	-.195 (-.480)	-.335 (-.500)
MSOHS	-.626 (-1.470)	-.968 (-1.410)
M79	-.922 ** (-1.970)	-1.050 (-1.370)
M06	-1.55** (-2.59)	-1.70* (-1.80)
MOCLER	1.79 (2.01)	-2.04 (-1.60)
MOSKIL	.864 (.960)	-2.09 (-1.55)
MOSEMI	.775 (1.000)	-1.99** (-2.04)
MOUNSK	.873 (1.230)	-2.11** (-2.25)
FEMALE	.419** (2.080)	.538* (1.740)
RURAL	.224 (.900)	.830 (1.210)
FHSG	.591* (1.640)	-.634 (-1.160)
FSOHS	.284 (.750)	-.330 (-.570)

Table 5.5 (cont.)

Independent Variable	Summer	Full Year
F79	.443	-.229
	(1.140)	(-.350)
F06	-.101	-.339
	(-.220)	(-.500)
FACLER	-.546	2.73*
	(-.880)	(1.89)
FASKIL	-.795*	1.32*
	(-1.780)	(1.87)
FASEMI	-1.100**	.422
	(-2.450)	(.630)
FAUNSK	-1.23**	.217
	(-2.65)	(.310)
BLACK	-1.55**	-2.04**
	(-3.16)	(-2.34)
MEXAM	-.724	-1.91**
	(-1.500)	(-2.21)
KIND	.770**	-.031
	(3.080)	(-.060)
BLKIND	-.578	.293
	(-1.200)	(.420)
HDSTRT	-.245	-.256
	(.940)	(-.480)
BLCKHS	.954**	1.18*
	(2.070)	(1.72)
MEXHS	.796	.519
	(1.270)	(.570)
CONSTANT	5.49**	7.38**
	(3.82)	(2.89)
R^2	.209	.444
N	623	206

Note: t-statistics are in parentheses below their coefficients.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

The regression coefficients in the tables relate the expected change in a child's test score when a variable is changed by one unit. To interpret the practical significance of the coefficients it is necessary to understand what a change of one point on the tests means in terms of cognitive development. In the ITPA manual Kirk et al. (1968) offer two interpretations to the total ITPA score. For children 5 to 7 years old the ITPA manual suggests that an increase of one point on the mean of the ten subtests (ITPAMN) is approximately equal to a gain of three months in the psycholinguistic age of the child. The ITPA manual also offers Stanford-Binet mental age estimates for the ITPA mean score; a gain of one point on the ITPA mean score is also approximately equal to a gain of three months of mental age. For children 5 to 7 years old a gain of three months in mental age is equivalent to an increase in IQ of four to five points. Thus, we can interpret the coefficients for the regressions using the ITPA mean score as the dependent variable in terms of gains in psycholinguistic age, mental age, and IQ. The MRT manual does not provide age or IQ equivalents, but it does rank the MRT scores by percentiles. For entering first grade children a gain of one point in the mean MRT score is approximately equal to a gain of twelve percentiles.[4] Because the regressions with ITPAMN as the dependent variable are easier to interpret, we shall confine our discussion to the results in Table 5.4.

The coefficients for the number of children in the family (CHILD) are both negative in Table 5.4, as expected, and the coefficient is significantly different from zero at the 5 percent level for the full-year regression. Using the conversion factors discussed above, the coefficient of -.36 for the full-year sample is equivalent to an expected decline of about one month of psycholinguistic age or mental age for each additional child in the family; this is also roughly equivalent to a decline of 1.5 IQ points. The coefficient for total family income (INCOME) is negative and insignificant for the summer sample, but positive and significantly different from zero for the full-year sample; the coefficient for the full-year sample indicates that psycholinguistic age will be increased by three months for each increment of $2,500 of income. Although we anticipated that the coefficient for income might be small, we have no good explanation for the significant difference in the coefficients for the two samples. The coefficients for the child's age (AGE) are positive and statistically significant, but the coefficient for the full-year sample is almost twice as large as for the summer sample.

Because we have included the information for education and occupation of the parents in sets of dummy variables, the coefficients within each set must be examined together to interpret the coefficients. The coefficients of the variables for education of the mother are all negative and the absolute values of the coefficients increase as we proceed from higher to lower levels of education for the summer sample; the same pattern is followed for the full-year sample except that the coefficient for mothers having less than seven years of education is larger than the two adjacent higher categories (M79 and MSOHS). Although the coefficients follow the expected pattern, the group of coefficients is not statistically significant at the 10 percent level for either sample. The coefficients for mother's occupational status do not follow the predicted pattern. For the summer sample all of the coefficients in the set are positive, and the coefficient for unskilled mothers (MOUNSK) is greater than the coefficient for semiskilled mothers (MOSEMI). These aberrations can perhaps be explained by the sample truncation problem and the coding scheme for unskilled mothers, but the pattern of the coefficients for this group is much different for the full-year sample where all of the coefficients are negative and the coefficient for clerical workers (MOCLER) is the smallest. It is apparent that the available information on mother's occupation does not fit into the educational production function in a stable or consistent pattern. Neither do the coefficients for father's education and occupation consistently conform to our expectations. The coefficients for father's education behave as predicted for the summer sample where only the variable for high school graduates (FHSG) has a positive coefficient and the coefficients decline as we move from higher to lower categories, but the coefficients are not statistically significant individually or together. For the full-year sample all of the coefficients are negative, but the ranking follows no pattern. The variables for father's occupation all have negative coefficients for the summer sample and they increase in size as we go from higher to lower categories. For the full-year sample the first two categories (FACLER and FASKIL) have positive coefficients and the coefficient for unskilled workers is less negative than the coefficient for semiskilled workers. It is apparent that the data do not strongly support our expectations concerning the effects of parental socioeconomic status, especially for the full-year sample.

The variables for ethnicity (BLACK and MEXAM) had the anticipated negative coefficients for the summer sample, but the coefficient for blacks was positive in the full-year sample. In the summer sample the coefficient for Mexican Americans is significant at the 10 percent level and the coefficient for blacks is significant at the 5 percent level; neither coefficient is significantly different from zero in the full-year sample. It is somewhat surprising that the Mexican American children scored higher than comparable blacks in the summer sample (but the difference is not statistically significant). Because of language problems, we would expect the Mexican American children to perform at lower levels.[5]

Head Start and kindergarten are the two most important policy variables that affect cognitive development. The coefficient of kindergarten for whites and Mexican Americans (which we constrained to be the same because there were too few Mexican Americans) is positive in both the summer and full-year samples, but there is a great difference in the magnitude of the coefficients. The coefficient for the summer sample is .86 which is roughly equal to 2.7 months of psycholinguistic age and 3.6 IQ points; for the full-year sample, however, the coefficient is 3.5 which is equivalent to almost 11 months of psycholinguistic age and 20 IQ points. Judging from our other results we suspect that the true effect is somewhere between these two values. To determine the effect of kindergarten for blacks we must add the values of the coefficients for kindergarten and for the black-kindergarten interaction dummy variable (BLKIND). For the summer sample the net effect of kindergarten for blacks is therefore .27, and for the full-year sample it is -.09. Unless there is some selectivity bias that operates differently for the two races it appears that kindergarten is much more effective for whites. Head Start has a negative, not statistically significant coefficient for white children in both samples. The Head Start-ethnicity interaction variables are positive in both samples and the black-Head Start interaction is significant at the 5 percent level for the summer sample. The net effect of Head Start for blacks is 1.6 points on the ITPA for the summer sample which is equivalent to 4.8 months of psycholinguistic age and an IQ gain of about 8 points. For the full-year sample the net effect of Head Start for blacks is only .82 ITPA points. It is somewhat surprising to find a smaller effect for the full-year program, and we suspect that this difference may be spurious. The Mexican American-Head Start interaction variable had a coefficient of .08 for the full-year sample so the net effect of Head Start is -1.14 for this group; for the summer sample the net effect of Head Start for Mexican Americans is 1.17 ITPA points, which is approximately equal to 3 months of psycholinguistic age and 6 IQ points. It is interesting to note that for the summer sample Head Start reduces the black-white IQ differential from 10 points to only 2 points.

To summarize briefly the regressions in Table 5.4, we have found that Head Start appeared to be very effective for blacks in the summer sample and somewhat effective for blacks in the full-year sample, but it appeared to be ineffective for whites in both samples. Kindergarten, on the other hand, is effective for whites but not for blacks. Some of the other independent variables had the expected sign, but there are many exceptions.

When the mean for MRT scores is used as the dependent variable the regressions are similar in form to those where ITPAMN is used. The only notable difference to be found between Tables 5.4 and 5.5 is that the coefficient for females is positive and significant when the mean for MRT is the dependent variable but is insignificant when the mean for ITPA is used.

4. Summary of Head Start and Kindergarten Effects for All Samples

The ITPA was designed so that it could be used for children up to 10 years old, and the Westinghouse researchers administered the test to children in all three grades. Table 5.6 summarizes the effects of Head Start and kindergarten for blacks and whites for all ten subsamples where the mean for ITPA scores is employed as the dependent variable; the complete regression results are reported in the Appendix. Although we are primarily interested in determining the initial impact of Head Start, it is also important to discover if there is a permanent effect or if it decays or increases in subsequent years. As we have noted in Chapter 4, only a longitudinal study is truly appropriate for measuring the decay of the Head Start effect. Since their inception the Head Start programs have changed--hopefully improved--and a smaller effect for the second and third grade samples may be due to this rather than to any actual decay. In addition, the selection process may have changed over the three-year period, perhaps leading to different biases.

For three of the four first grade samples Head Start has a positive effect for black children, and these effects are equivalent to four to ten IQ points; only for the summer, mother only sample is the effect

Table 5.6

Summary of Effects of Head Start and Kindergarten for All Regressions with ITPAMN as the Dependent Variable

Grade	Type of Program	Parents Present	Observations	Head Start for Whites	Head Start for Blacks	Kindergarten for Whites	Kindergarten for Blacks
1	full year	both	206	-1.22	.82	3.47**	-.09**
1	full year	mother	67	3.86*	1.33	5.21*	2.09
1	summer	both	623	-.36	1.63**	.86**	.27
1	summer	mother	143	2.62**	-.82	1.22	.42
2	full year	both	218	-.52	.08	3.57**	.77**
2	full year	mother	75	1.18	1.70	1.98	1.29
2	summer	both	635	-.48	-.57	1.46**	-.10**
2	summer	mother	134	1.11	.51	1.94**	1.60
3	summer	both	426	-.02	.58	1.97**	1.70
3	summer	mother	134	1.52	-.48	1.23	.52

* Statistically significant at the 10 percent level for white coefficients.
** Statistically significant at the 5 percent level for white coefficients.
* Statistically significant at the 10 percent level for black interaction coefficients.
** Statistically significant at the 5 percent level for black interaction coefficients.

negative (and then it is almost zero). Thus, the immediate impact of Head Start for black children is quite favorable. For the white children in the first grade, Head Start seems to be effective only when the father is absent from the home. For the second and third grade samples the effect of Head Start is not great for black children except for the grade 2 summer sample where only the mother is present. The white children show the same pattern for the second and third grades as for the first grade--Head Start has a fairly high effect (greater than one point on ITPAMN) when only the mother is present and a very small, negative effect when both parents are present. Thus, Head Start appears to be effective for white children from mother-headed families and for all black children.

Kindergarten consistently shows a strong, positive effect for white children in all of the samples. It is interesting to note that the kindergarten coefficients for white children remain large and significant even through the third grade. The effect of kindergarten for blacks is less than it is for whites in all ten samples, and is slightly negative for two samples. There is no clear trend for the kindergarten effects for blacks, and the coefficients vary so much that it is impossible to determine how effective kindergarten is for black children.

5. Techniques for Measuring Head Start and Control Pretreatment Differences

As we have pointed out in previous chapters, a great deal of the controversy about the Westinghouse study concerns the question of whether or not the experimental and control groups differ on their pretreatment levels of cognitive development. In Chapter 2 we demonstrated that under certain conditions a difference favoring the control group would bias the estimated effect of Head Start toward zero. Because there are no pretest scores available, the best one can do is to examine the vector of independent variables to determine if one group appears more advantaged previous to Head Start enrollment.

There are several approaches that can be used to determine if the groups differ in their pretreatment socioeconomic status. Perhaps the simplest is to perform an analysis of variance for each trait of interest (such as income, family size, and education of the mother) to see if the experimentals and controls differ on each of these variables. The technique we have used is called discriminant analysis, and it is used to test the hypothesis that the two groups differ significantly on the entire set of independent variables. Thus by using discriminant analysis we need only test one hypothesis per sample.

The discriminant analysis can be carried out in the regression framework by regressing the dummy variable for Head Start on the same set of independent variables included in the educational production function. In the limiting case where there is no relationship between the independent variables and group membership all the coefficients would be zero and the constant would be 0.5 (assuming equal sample size in the two groups). Because we are dealing with a 0-1 dependent variable, the fitted values of the dependent variable can be interpreted as the probability that a particular observation will be in Head Start rather than the control group.[6] The regression coefficients can then be interpreted as the change in the probability of membership in the Head

Start group for a unit change in the independent variables. Thus a positive coefficient for a variable indicates that the higher an observation's score is on that variable the more likely the observation is to be in the Head Start group. The F-statistic for the test that the entire set of coefficients (excluding the constant) are equal to zero is equivalent to the test of the hypothesis that the means of each variable are the same in the two groups.

The results of the discriminant analyses for the first grade, both parents present samples are found in Table 5.7. The F-statistic is below the critical value for the full-year sample and the hypothesis of no difference cannot be rejected at the 5 percent level. For the summer sample the means are significantly different at the 1 percent level, so we must examine the coefficients to determine what direction the differences take. The coefficient for number of children in the family is .02 which can be interpreted that the probability of a child in the sample being in the Head Start group rather than the control group is increased by .02 for each sibling he has. The coefficient for income is negative and statistically significant and indicates that for every $1,000 of additional family income a child's probability of being in Head Start is reduced by .02. The coefficients for income and family size indicate that on these criteria the control group is more advantaged. Within a set of dummy variables, such as mother's education, the coefficient of a variable gives the change in probability for being in that group rather than the base group. For example, the probability of a child being in Head Start is reduced by .12 if the child's mother

Table 5.7

Discriminant Regression Results for Grade 1, Both Parents Samples, with Head Start as the Dependent Variable, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	.024**	-.002
	(2.380)	(-.130)
INCOME	-.00002**	-.00003
	(-2.12000)	(-1.45000)
AGE	-.056	-.019
	(-1.420)	(-.230)
MHSG	-.116	.195
	(-1.430)	(1.210)
MSOHS	.052	.131
	(.620)	(.800)
M79	-.092	.032
	(-.990)	(.170)
M06	-.041	-.071
	(-.350)	(-.320)
MOCLER	-.277	.018
	(-1.580)	(.060)
MOSKIL	-.075	.134
	(-.420)	(.410)
MOSEMI	-.045	.090
	(-.300)	(.380)
MOUNSK	-.220	.099
	(-1.570)	(.440)
FEMALE	-.001	-.021
	(-.040)	(-.280)
RURAL	.060	.088
	(1.220)	(.580)
FHSG	-.058	.289
	(-.810)	(2.210)
FSOHS	-.083	.355
	(-1.120)	(2.570)
F79	.049	.334
	(.640)	(2.140)

Table 5.7 (cont.)

Independent Variable	Summer	Full Year
F06	-.004 (-.050)	.334** (2.070)
FACLER	.171 (1.390)	-.649* (-1.880)
FASKIL	.112 (1.270)	-.226 (-1.340)
FASEMI	.135 (1.530)	-.241 (-1.440)
FAUNSK	.220** (2.410)	-.134 (-.780)
KIND	.058 (1.320)	.027 (.320)
BLACK	-.021 (-.410)	.017 (.120)
MEXAM	.027 (.390)	.077 (.480)
CONSTANT	.894** (3.170)	.433 (.730)
R^2	.102	.100
F-statistic	2.820	.842
N	623	206

Note: t-statistics are in parentheses below their coefficients. For the summer sample F-statistic numerator degrees of freedom = 24; denominator degrees of freedom = 598. For the full-year sample F-statistic numerator degrees of freedom = 24; denominator degrees of freedom = 181.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

is a high school graduate rather than if she had some college. It appears that for the summer sample the control group does come from a more advantaged background, but the discriminant analysis does not tell us how much this difference will make in terms of cognitive development.

The coefficient for kindergarten is subject to the same potential bias problem as Head Start, and Table 5.8 contains results from the discriminant analysis with kindergarten attendance as the dependent variable for the same samples as those used in Table 5.7 There is a much greater chance that the kindergarten and nonkindergarten children differ in background because no attempt was made in the Westinghouse study to equate these two groups. The F-statistics are significant for both the summer and full-year samples at the 1 percent level. For both samples it appears that the children who have attended kindergarten are from higher socioeconomic backgrounds on the criteria of family income and education of the father.

Although discriminant analysis is one way to determine if there are significant differences between groups, it has several weaknesses. If some differences favor the control group and others favor the Head Start children, we have no method by which to include these differences. Even if all of the evidence indicates that one group comes from a better background, discriminant analysis does not give any information about what the magnitude of these differences will be in terms of cognitive development. To get a better idea of the differences in cognitive development that would have been observed if there

Table 5.8

Discriminant Regression Results for Grade 1, Both Parents Samples, with Kindergarten as the Dependent Variable, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	.026* (2.710)	-.007 (-.440)
INCOME	.00004** (5.19000)	.00003** (1.99000)
AGE	.046 (1.240)	-.055 (-.750)
MHSG	.030 (.390)	-.332** (-2.360)
MSOHS	-.010 (-.120)	-.185 (-1.280)
M79	-.096 (-1.100)	-.344** (-2.150)
M06	-.074 (-.670)	-.541** (-2.760)
MOCLER	.236 (1.440)	.320 (1.190)
MOSKIL	.451** (2.720)	.019 (.070)
MOSEMI	.279** (1.960)	.252 (1.200)
MOUNSK	.256** (1.970)	.202 (1.010)
FEMALE	-.009 (-.250)	.031 (.460)
RURAL	-.152** (-3.330)	-.598** (-4.710)
FHSG	-.046 (-.690)	-.051 (-.440)
FSOHS	-.108 (-1.550)	-.053 (-.430)

Table 5.8 (cont.)

Independent Variable	Summer	Full Year
F79	-.142**	-.085
	(-1.990)	(-.620)
F06	-.196**	-.191
	(-2.350)	(-1.340)
FACLER	.153	.252
	(1.330)	(.830)
FASKIL	.010	.048
	(.120)	(.320)
FASEMI	.017	.033
	(.210)	(.220)
FAUNSK	-.034	.024
	(-.400)	(.160)
BLACK	.130**	-.326**
	(2.730)	(-2.800)
MEXAM	.302**	-.263*
	(4.840)	(-1.880)
CONSTANT	-.151	1.22*
	(-.570)	(2.34)
R^2	.166	.269
F-statistic	5.200	2.920
N	623	206

Note: t-statistics are in parentheses below their coefficients. For the summer sample F-statistic numerator degrees of freedom = 23; denominator degrees of freedom = 599. For the full-year sample F-statistic numerator degrees of freedom = 23; denominator degrees of freedom = 182.

* Statistically significant at the 10 percent level.

** Statistically significant at the 5 percent level.

had been no treatment, we have developed a two-stage procedure as an alternative to the discriminant analysis.[7] The first stage of the procedure is to regress the cognitive measure (such as ITPAMN) on all of the independent variables for observations in the control group. This yields an estimate of the educational production function without Head Start. This fitted function is then used to form an imputed test score for observations in both groups; the imputed score is formed by the formula $\mathrm{YHAT} = \sum_{i=1}^{k} \beta_i X_i$ for each of the observations. In the second stage of the procedure the variable YHAT is regressed on the dummy variable for Head Start. The coefficient for the Head Start variable then relates the difference in the average test scores for the two groups.

Because this procedure is expensive to carry out we have only applied it to the first grade summer and full-year samples where both parents are present. For the summer sample when ITPAMN is used as the cognitive measure in the first stage, the second stage regression equation is:

$$\begin{array}{llll} \mathrm{YHAT} = & 19.300 & -.764 & \mathrm{HDSTRT} \\ & (183.508) & (-5.047) & \end{array} \qquad (1)$$

$N = 623$; $R^2 = .0394$. Values in parentheses are t-statistics. To clarify the meaning of equation (1) we shall repeat the steps of the procedure used to arrive at this regression. The first step in the procedure is that the ITPA mean score was regressed on the set of socioeconomic and demographic variables (excluding the Head Start and Head Start-ethnicity interaction variables) for the observations in the control group. The constant and regression coefficients were then assumed to represent the parameters of the educational production function for children who have not had Head Start. Imputed test scores (YHAT) for the control children and Head Start children were then formed by multiplying the coefficients by the appropriate values of the independent variables. For the control group the mean of the imputed scores is equal to the mean of the actual scores on the ITPA mean, but this is not necessarily the case for the Head Start children. Then the imputed scores (YHAT) were regressed on the dummy variable for Head Start (HDSTRT). The constant in this regression gives the mean of the imputed (and actual) ITPA scores for the control group; the mean of the imputed scores for Head Start group is equal to the sum of the constant and the coefficient of Head Start. For equation (1) the mean of the imputed scores for the control group is 19.30, and the mean of the imputed scores for the Head Start group is 18.54. The difference in the imputed means, the coefficient of Head Start, is -.76 points on the ITPA and this difference is significant at the 1 percent level. This difference is equivalent to about 2.3 months of psycholinguistic age or roughly 3.8 IQ points. Thus, for the present sample we find that the control group appears to be more able to the extent of 3.8 IQ points, and that this difference is statistically significant. When MRTMN is used as the dependent variable in the first stage, the regression of YHAT on HDSTRT is:

$$\begin{array}{lll} \text{YHAT} = & 9.306 & -.421\ \text{HDSTRT} \\ & (116.218) & (-3.737) \end{array} \qquad (2)$$

$$N = 623;\ R^2 = .0220$$

For the first grade full-year sample with both parents present, the second-stage regression when ITPAMN is used in the first stage is:

$$\begin{array}{lll} \text{YHAT} = 19.655 & -.688\ \text{HDSTRT} & \\ \quad (78.650) & (-1.958) & (3) \end{array}$$

$N = 206;\ R^2 = .0185$

Where MRTMN is used in the first stage as the dependent variable, the regression of YHAT on HDSTRT for this sample is:

$$\begin{array}{lll} \text{YHAT} = 8.805 & -.438\ \text{HDSTRT} & \\ \quad (46.304) & (-1.630) & (4) \end{array}$$

$N = 206;\ R^2 = .0129$

This alternative to the discriminant analysis thus gives information on how much lower cognitive development the Head Start group would have compared to the control group if they had not participated in Head Start. This procedure does not tell us whether or not our previous estimates of Head Start are biased, because we do not know which selection procedure was used in assigning children to Head Start.

Our inability to resolve the bias issue illustrates the major weakness of ex post facto evaluations. What is clearly needed is detailed information on the selection procedures used by the Head Start centers. The Head Start center directors should have been required to give detailed information on the procedures utilized. The Parent Interview Questionnaire should have sought information about how the Head Start children were recruited for the program and the reasons why the control children were not enrolled. If this additional information were available, we would be in a better position to resolve the bias issue.

6. Supplementary Models

We have made several attempts to expand the regression model we have worked with to include more than the socioeconomic and demographic information available. The extensions to the basic models that we have tried are adding SES-Head Start interaction terms for all socioeconomic measures, adding dummy variables for neighborhoods and individual centers, adding Head Start center characteristics to the regression model, and stratifying the samples by the employment status of the mother. The extensions have only been carried out for the first grade, summer and full-year samples with both parents present mainly because we have found that our basic results were little affected by these extensions.

By including interaction variables for Head Start and the socioeconomic and demographic variables already in the model, we allow the effect of Head Start to vary for children from different backgrounds. The basic model already includes interaction terms for Head Start and ethnicity, and these interaction variables have significant coefficients in many of the samples. The interaction variables are formed by multiplying the dummy variable for Head Start by each of the other independent variables. If, for example, Head Start is more effective for children from families with high incomes then the coefficient for the income-Head Start interaction term will be positive; if the effect of Head Start is independent of family income then the coefficient of that interaction variable will be zero. Since we had already decided that the ethnicity-Head Start interactions should be included in the model we tested the hypothesis that all of the other Head Start interaction variables had coefficients of zero. For each of the two samples we ran regressions using the ITPA mean and the MRT mean as the dependent variables. The null hypothesis of no significant interactions was not rejected at the 10 percent level for all four regressions. For the full-year sample the test statistics are $F_{21,\ 156} = .760$ and $F_{21,\ 156} = 1.20$ for the regressions using the ITPA mean and MRT mean, respectively. For the summer sample the corresponding statistics are

$F_{22,\ 572} = 1.22$ and $F_{22,\ 572} = .77$. Because of these findings we have rejected the hypothesis that the effect of Head Start is dependent upon the socioeconomic background of the family.[8]

The analysis of covariance model used in the Westinghouse model included dummy variables for neighborhood; a dummy variable was established for each target area and the neighborhood was thus allowed to have an effect on the test score. We have extended our basic model by adding neighborhood variables, and we have extended it further by adding dummy variables for individual Head Start centers. The Head Start dummy variables were added for all but one center, and the dummy variable for Head Start was retained in the model; thus the F-test on the significance of the center variables tests the null hypothesis that all of the Head Start centers have the same effect as the base center (which was selected randomly). For these regressions we included only those observations from neighborhoods where there were at least four Head Start and four control observations.

For the full-year centers we found the set of neighborhood coefficients to be significantly different than zero at the 1 percent level when the ITPA mean is used as the dependent variable, but that the coefficients for the set were not significant when the MRT mean is used. The test statistics are $F_{8,\ 72} = 4.24$ and $F_{8,\ 72} = .580$, respectively. We then tested the significance of the Head Start center dummy variables when they are added to the equation containing the socioeconomic and demographic variables and the neighborhood variables. The set of center variables was not significant when either the ITPA mean or the MRT mean was used as the dependent variable; the corresponding test statistics are $F_{8, 64} = 1.01$ and $F_{8, 64} = .660$.

The same tests were carried out for the first grade, summer, both parents present sample with similar results. The set of neighborhood coefficients is significant when either the ITPA mean or the MRT mean is used as the dependent variable; the F-statistics for these tests are $F_{32, 327} = 3.83$ and $F_{32, 327} = 3.70$. Once again, the Head Start center dummy variables are insignificant. The test statistics for these sets are $F_{33, 294} = 1.10$ and $F_{33, 294} = 1.20$ for the ITPA mean and MRT mean used as the dependent variables; neither of these values is significant at the 10 percent level.

The implication of the finding that the center effects are all about the same is that either there is not a great deal of variation between centers or that the variation that exists does not affect the centers' impact on the children's cognitive development. An examination of the responses to the Head Start Official's Interview Questionnaire (HSIQ) leads us to agree with the Westinghouse report that "In sum, the data furnished by the HSIQ leads to the conclusion that there is a large common core of similar structures, objectives, and curriculum experiences in all Head Start programs" (p. 120). In spite of the homogeneity of the individual centers, we felt that to fully exploit the available data an attempt should be made to determine if certain center characteristics lead to more successful centers. From the HSIQ the variables that might affect cognitive development were selected; a description of the Head Start center characteristic variables appears in Table 5.9. Some of the variables of interest are measured in

Table 5.9

Description of Head Start Center Characteristic Variables

Variable	Description
PUBOP	1 if center is operated by public school; 0 otherwise
PROP	1 if center is operated by private or parochial school, religious related organization, or private nonprofit group; 0 otherwise
COLOP	1 if center is operated by a college; 0 otherwise
CAPOP	1 if center is operated by a community action agency; 0 otherwise
AGEEN	median age of children at entry of Head Start
CHITEA	number of children per teacher at the center
CHISTA	number of children per teacher and aides at the center
CENSIZ	total number of children at the center
PASTCH	ratio of parents serving as volunteers and paid staff to children in the center
HOMEVI	average number of home visits per child per month
STVIS	1 if the center stresses improvement in visual and hearing sensitivity, muscle development, and coordination very often; 0 otherwise
STGRAM	1 if the center stresses improvement in grammar, vocabulary, communication skills, and understanding very often; 0 otherwise
STTIME	1 if the center stresses improvement in the concepts of time, color, size, and functional relations very often; 0 otherwise
STINT	1 if the center stresses improvement in interest in learning and achievement very often; 0 otherwise
SCIIM	1 if the center considers science experience to be of very great importance to its curriculum; 0 otherwise
MATHIM	1 if mathematics is of very great importance to the center's curriculum; 0 otherwise
LANGIM	1 if language development is of very great importance to the center's curriculum; 0 otherwise

Table 5.9 (cont.)

Variable	Description
PROBIM	1 if problem solving is of very great importance to the center's curriculum; 0 otherwise
MONT	1 if the center uses Montessori methods; 0 otherwise
DACARE	1 if the center is basically a day care center; 0 otherwise
RESEN	1 if the center has a responsive environment; 0 otherwise
DRILL	1 if the center uses structured drills; 0 otherwise
ENRICH	1 if the center provides environmental enrichment; 0 otherwise

relative terms which makes it difficult to compare responses across centers. For example, the variables relating to the objectives of the center (STVIS, STGRAM, STTIME, and STINT) had the center director respond whether a concept was stressed not at all, not very often, often, or very often. Virtually all of the centers responded that all of the concepts were stressed often or very often, so to discriminate in the model we defined the dummy variables for this set on the basis that if a concept was stressed very often the variable was set equal to one, and it was set equal to zero otherwise. Unfortunately, what one official considers "often" another may consider "very often" so these variables may not have a great deal of meaning. The same problem arises for the variables on center curriculum (SCIIM, MATHIM, LANGIM, and PROBIM) where most of the responses were either "of very great importance" or "of some importance." Another problem with the HSIQ is that there is no information about the number of hours per day that the children spend at the center. Given the homogeneity of the responses and imprecise form of the questions on the HSIQ and our finding that all centers appear to be equally effective, we did not expect to learn much from the analysis of the center characteristics.

The analysis of center characteristics was carried out for the first grade, summer and full-year, both parents present samples. Only the children from the Head Start group were used, and the center characteristics were entered into the regression equation with the socioeconomic and demographic variables. Both the ITPA mean and MRT mean were employed as the dependent variable, and the regressions are presented in Table 5.10 and Table 5.11.

Table 5.10

Effects of Individual and Center Characteristics for Grade 1, Both Parents Present, on Child's ITPA Score, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	-.007 (-.056)	-.646** (-2.834)
INCOME	-.0001 (-1.2892)	.393* (1.643)
AGE	.650 (1.233)	-.337 (-.319)
MHSG	-.504 (-.513)	.416** (2.246)
MSOHS	-1.14 (-1.16)	.411 (1.968)
M79	-.614 (-.580)	.460** (2.132)
M06	-2.89** (-2.31)	.267 (1.153)
MOCLER	-1.19 (-.51)	.531 (.713)
MOSKIL	-.431 (-.206)	----
MOSEMI	-3.48** (-1.96)	.472 (1.142)
MOUNSK	-2.75* (-1.68)	.127 (.306)
FEMALE	-.017 (-.038)	.634 (.845)
RURAL	.121 (.167)	.104 (.728)
FHSG	1.10 (1.32)	-.992 (-.511)
FSOHS	-.161 (-.185)	.107 (.505)

Table 5.10 (cont.)

Independent Variable	Summer	Full Year
F79	.208 (.233)	-.998 (-.442)
F06	-.289 (-.293)	.168 (.715)
FACLER	1.84 (1.20)	----
FASKIL	.169 (.146)	.224 (1.089)
FASEMI	.336 (.291)	.181 (.999)
FAUNSK	-.265 (-.230)	.149 (.771)
BLACK	-.968 (-.921)	.152** (2.641)
MEXAM	.582 (.641)	.590 (1.355)
KIND	1.13 (1.46)	.281* (1.770)
BLKIND	-.498 (-.415)	-.831 (-.178)
PROP	-.055 (-.043)	----
CAPOP	-.419 (-.689)	.229 (.798)
AGEEN	-.846 (-1.462)	.294** (1.731)
CHITEA	-.154 (-1.065)	-.822 (-.545)
CHISTA	.248** (2.404)	-.204 (-.153)
CENSIZ	-.007 (-1.569)	-.736 (-.120)

Table 5.10 (cont.)

Independent Variable	Summer	Full Year
PASTCH	.354	-.193
	(.749)	(-.486)
HOMEVI	-.582	-.247
	(-.619)	(-1.304)
STVIS	-1.46 **	.676
	(-2.02)	(1.353)
STGRAM	1.08	-.654
	(1.60)	(-.997)
STTIME	-.657	-.899
	(-.901)	(-1.089)
STINT	1.46 *	-1.968
	(1.79)	(-.411)
SCIIM	.253	-.393
	(.377)	(-.064)
MATHIM	-1.00	-.350
	(-1.53)	(-.703)
LANGIM	-2.16	-.169
	(-1.88)	(-.233)
PROBIM	-.682	.899
	(-1.077)	(1.029)
MONT	.857	-.284
	(1.489)	(-.954)
DACARE	-.660	----
	(-.635)	
RESEN	.980	.480
	(.903)	(.952)
DRILL	-1.36 *	
	(-1.76)	
ENRICH	.175	
	(.196)	
CONSTANT	26.893 **	11.558
	(4.515)	(.476)

Table 5.10 (cont.)

R^2	.3212	.8186
N	262	70

Note: t-statistics are in parentheses below their coefficients. A dashed line indicates that the variable was not entered because of zero variance. For the full-year sample no coefficients or t-statistics are reported for the drill or enrichment program because of multicollinearity.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

Table 5.11

Effects of Individual and Center Characteristics for Grade 1, Both Parents Present, on Child's MRT Score, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	-.056 (-.646)	-.303* (-1.869)
INCOME	-.00007 (-.93200)	.239 (1.407)
AGE	.532 (1.371)	.894 (1.192)
MHSG	-.367 (-.509)	.506 (.384)
MSOHS	-1.08 (-1.50)	-.238 (-.160)
M79	-1.10 (-1.41)	-.356 (-.023)
M06	-1.80** (-1.95)	-.137 (-.834)
MOCLER	.221 (.129)	-.759** (-2.012)
MOSKIL	-.294 (-.191)	----
MOSEMI	-1.37 (-1.05)	.709
MOUNSK	-1.21 (-1.01)	-.833** (-2.832)
FEMALE	.470 (1.412)	.106** (1.979)
RURAL	.483 (.903)	.108 (.001)
FHSG	.906 (1.480)	.225 (.163)
FSOHS	.633 (.990)	.187 (1.241)

Table 5.11 (cont.)

Independent Variable	Summer	Full Year
F79	.891 (1.361)	.646 (.403)
F06	.685 (.945)	.879 (.528)
FACLER	1.30 (1.16)	----
FASKIL	.094 (.111)	.214 (.146)
FASEMI	-.248 (-.293)	.234 (.182)
FAUNSK	-.528 (-.622)	-.512 (-.374)
BLACK	-2.13 ** (-2.76)	.960 (.235)
MEXAM	.578 (.867)	.119 (.385)
KIND	1.28 ** (2.26)	.164 (.146)
BLKIND	.938 (1.066)	.453 (.136)
PROP	.946 (.010)	----
CAPOP	.912 ** (2.043)	.813 (.399)
AGEEN	-.361 (-.850)	.118 (.977)
CHITEA	-.131 (-1.237)	.674 (.629)
CHISTA	.170 ** (2.253)	-.710 (-.751)
CENSIZ	.003 (.766)	-.266 (-.610)

Table 5.11 (cont.)

Independent Variable	Summer	Full Year
PASTCH	.304 (.875)	-.282 (-.998)
HOMEVI	-.715 (-1.036)	-.256* (-1.902)
STVIS	-.600 (-1.133)	.102** (2.861)
STGRAM	.753 (1.514)	-.860* (-1.845)
STTIME	-.279 (-.521)	.281 (.480)
STINT	.602 (1.002)	-.478 (-1.405)
SCIIM	.109 (.222)	-.460 (-1.059)
MATHIM	-.155 (-.323)	.171 (.483)
LANGIM	-1.37 (-1.63)	-.402 (-.777)
PROBIM	-.387 (-.832)	-.581 (-.937)
MONT	1.06** (2.50)	-.199 (-.940)
DACARE	.641 (.840)	----
RESEN	.903 (1.132)	.111 (.310)
DRILL	-.337 (.595)	
ENRICH	.476 (.728)	
CONSTANT	8.241* (1.885)	11.896 (.689)

Table 5.11 (cont.)

R^2	.311	.119
N	262	70

Note: t-statistics are in parentheses below their coefficients. A dashed line indicates that the variable was not entered because of zero variance. For the full-year sample no coefficients or t-statistics are reported for drill or enrichment programs because of multicollinearity.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

For some center characteristics we had expectations about the sign of the coefficients, but for others we had no a priori hypotheses and we hoped to learn if the characteristics affected the performance of the Head Start participants. It is possible that the agency sponsoring the program would affect center effectiveness, so two dummy variables (PROP and CAPOP) were included to test this. We expected the effectiveness of the center to decrease as the child-teacher ratio (CHITEA) and child-staff ratio (CHISTA) increase. We had no expectations about the effect of center size (CENSIZ), but we felt that it would be interesting to test the contention of Smith and Bissell that small centers are more effective. The variables for home visits (HOMEVI) and for the ratio of parent volunteers and paid staff to children (PASTCH) measure the interaction between the child's family and the Head Start program, and we expected the coefficients for the variables to have positive coefficients. We expected the coefficients for the variables indicating that cognitive and language skills are considered important by the program (STVIS, STGRAM, STTIME, STINT, SCIIM, MATHIM, LANGIM, and PROBIM) to have positive coefficients, as many past studies have found that preschool programs that stress cognitive goals generally achieve the greatest gains on cognitive and achievement tests. Regarding the general approach of the center, we expected the day care centers (DACARE) to produce a smaller effect than those that used drill (DRILL), Montessori methods (MONT), responsive environments (RESEN), or enrichment (ENRICH), but we had no strong rankings among the other four approaches; the variables for center approach were not categorized as being mutually exclusive, and many centers responded that they used more than one approach (such as

Montessori and responsive environment). Because the full-year programs had a longer period of time to implement their curricula, we suspected the coefficients for the objectives and curriculum of the center would be larger for the full-year sample than for the summer sample.

Although it was realized that the Westinghouse data is not well suited for an analysis of the effects of center characteristics, the results contained in Table 5.10 and Table 5.11 are so contradictory and inconclusive as to give virtually no useful information about the effects of center characteristics. For example, the coefficient for the child-teacher ratio (CHITEA) is negative as expected for both the full-year and summer samples when the ITPA mean is used as the dependent variable, but it is positive for the full-year sample when the MRT mean is used as the dependent variable; in none of the cases is the coefficient significantly different from zero. The coefficient for the child-staff ratio (CHISTA) is positive and significant at the 5 percent level for the summer sample when both cognitive measures are used as the dependent variable; we can think of no logical explanation for this finding. Perhaps the most perplexing finding is that over half of the coefficients for the variables measuring the importance of cognitive development in the Head Start program had negative coefficients--sometimes statistically significant. This might be explained in part by a reciprocal causation model (centers stressed the areas in which the children were weakest). We do not feel that this adequately explains our findings. We must conclude that the Westinghouse data is totally inappropriate for evaluating the differential effects of center characteristics, and that one should not make inferences regarding the efficacy of different approaches to the curriculum and organization of the centers from our analysis.

The final supplementary analysis carried out was performed to determine if there is an interaction between the employment status of the mother and the effectiveness of Head Start. For two-parent families we would expect Head Start to be more effective when both parents are employed because the child will not benefit from the presence of the mother in the home; hopefully, Head Start will provide the care and guidance that the child might otherwise not receive. Although this hypothesis holds for mother-headed families too, there is an additional complication to be considered. When the mother is the only parent present and she is not employed, the family must be deriving most of its income from welfare and other transfer payments; the children in these families are among the most disadvantaged in our sample, and it is likely that any Head Start-SES interaction effect (positive or negative) would show up for them.

There is currently a great deal of debate concerning whether or not mothers who are heads of households should be encouraged to work or remain at home with their young children. Many people feel that these mothers should be encouraged to work by providing subsidized child care (by Head Start programs or subsidized day care centers) or by making day care expenses tax deductable. Thus the employment status of the mother is subject to policy manipulation, and the results of our analysis can be used as evidence to determine if Head Start is a good alternative to the mother's presence in the home. (Of course we must note that our data is not ideal for such purposes since there may be differences between the working and nonworking mothers that we are unable to control for.)

For this supplementary analysis we used only children in the first grade because for the children in the second and third grades the data on the employment status of the mother are not likely to be reliable measures of her employment status when her child was eligible for Head Start. The data were stratified by type of Head Start program (summer and full-year), parents present (both and mother only), and employment status of the mother (employed or not employed). The analysis was not carried out for the full-year, mother-only subsamples because there were too few observations available. We have included the same socioeconomic and demographic independent variables in our analysis as we have used in the previous regressions except that the variables for occupation of the mother were omitted from the analysis for the subsamples where the mother was not employed. We have used both the ITPA mean and the MRT mean as dependent variables for these analyses. The regression coefficients for all the independent variables except Head Start are close to those found when we did not stratify by mother's employment status, and these coefficients are reported in Tables 5.4 and 5.5 of the text and A-20 through A-27 of the Appendix; the complete regressions are available from the author upon request. We have summarized the Head Start coefficients for these analyses in Table 5.12.

Table 5.12

Summary of Head Start Coefficients when Data is Stratified by Employment Status of the Mother

Type of Program	Parents Present	Employment Status of the Mother	Observations	Head Start for Whites	Head Start for Blacks
I. ITPA mean is used as the dependent variable					
summer	both	employed	196	-.524	1.632
summer	both	not employed	427	-.243	1.427**
summer	mother	employed	48	4.153	.487
summer	mother	not employed	95	2.353*	-.298
full-year	both	employed	74	-1.893	.335
full-year	both	not employed	132	-.741	.916
II. MRT mean is used as the dependent variable					
summer	both	employed	196	.184	1.077
summer	both	not employed	427	-.411	.442
summer	mother	employed	48	2.291	.667
summer	mother	not employed	95	.142	-.119
full-year	both	employed	74	-2.072	.981**
full-year	both	not employed	132	.216	.784

* Statistically significant at 10 percent level for white coefficients.

** Statistically significant at 5 percent level for white coefficients.

* Statistically significant at 10 percent level for black interaction coefficients.

** Statistically significant at 5 percent level for black interaction coefficients.

The data in Table 5.12 do not strongly support the hypothesis that Head Start is more effective for children of working mothers than for children of nonworking mothers. For white children there is an appreciable difference in the effects of Head Start when we stratify by employment status of the mother when the mother is the head of the family; Head Start has a higher coefficient for children of working mothers when either the ITPA mean or MRT mean is used as the dependent variable. For the other samples of white children either the effects of Head Start are greater when the mother does not work or the difference is small. For black children, Head Start was more effective when the mother works, but the differences are not great; when the ITPA mean is used as the dependent variable the greatest difference again occurs when the mother is the only parent present. It is unfortunate that there were too few observations for the full-year, mother-only sample to be stratified by employment status of the mother, for it would prove interesting to test the hypothesis for a program of longer duration. Our results support the hypothesis that Head Start programs can be especially useful for mother-headed families when the mother is employed, but we should bear in mind that our analyses are based on very small samples; clearly this is an area where additional research is needed.

7. Summary of the Appendix

The Appendix includes many of the results that were interesting, but not of primary importance. The first group of tables in the Appendix considers the use of the ITPA and MRT subtests as dependent variables for the regressions rather than using only the means. Table A-1 gives the correlations of all subtest and mean scores for the first grade, summer, both parents sample. All of the subtests are positively correlated, ranging in value from .13 to .60; thus the subtests are not so highly correlated as to be measuring the same skills and abilities. Tables A-2 through A-19 contain the regressions of the subtests and the mean scores on the socioeconomic and demographic variables for this same sample. There are no great differences between the regressions using the ITPA mean and MRT mean as the dependent variable and those using the subtests; the coefficients for Head Start when the means are used are approximately equal to the mean of the subtest coefficients.

The next group of tables, Table A-20 through Table A-39 contain the regressions for all samples where the ITPA mean and the achievement test mean (MRT or SAT depending on the grade level) are used as the dependent variables. A summary of the coefficients of Head Start and kindergarten when the ITPA mean is used as the dependent variable was presented in Table 5.6, and the effects are generally of the same sign when the MRT, SAT2, or SAT3 is used as the dependent variable. There are once again many cases where the coefficients of some of the socioeconomic variables (such as income and children) do not have the expected sign.

The discriminant regressions with Head Start as the dependent variable are contained in Table A-40 through Table A-49. The test of differences in the vector of independent variables is significant for only three of the ten samples at the 5 percent level. Only for the first grade, summer, both parents sample is it apparent that the differences favor the control group. The discriminant analyses with kindergarten as the dependent variable are in Table A-50 through Table A-59. The F-statistics for all ten subsamples are highly significant, and the coefficients indicate that the children who attended kindergarten come on average from significantly more advantaged backgrounds. It is thus very likely that the coefficients for kindergarten attendance are inflated due to these differences.

8, Comparison of the Reanalysis to the Westinghouse Findings

Because we have analyzed the data differently from the Westinghouse-Ohio University study it is not surprising that our results are somewhat different. The major changes we have made are that we have stratified the data differently, we have used individual rather than grouped data, and we have included many more socioeconomic and demographic variables as independent variables.

The Westinghouse study found the summer Head Start programs to be totally ineffective. They found no significant effect on the ITPA for all three grades. If we had stratified the data in the same manner it is likely that we would have found similar results. However, we allowed Head Start to have different effects for different ethnic groups and for mother-headed families. For white children from two-parent families the coefficient for Head Start is negative but not statistically significant for all three grades for the summer samples where the ITPA mean is used as the dependent variable. For white children from mother-headed families, however, the effects of Head Start on the ITPA mean are 2.6, 1.1, and 1.5 for the first, second, and third grades, respectively. Although only the coefficient for the second grade sample is significant at the 5 percent level, the magnitudes of the coefficients are the equivalent of over three months of psycholinguistic or mental age. This is surely an impressive effect for a summer program. The effects of summer Head Start for the black children were not impressive. Only for the first grade sample of blacks with both parents is the effect of summer Head Start greater than one point; for that sample the effect is 1.6 points on the ITPA mean and is significant at the 5 percent level. Since white children with both parents present comprise over half of the observations at each grade level, it is not surprising that the Westinghouse researchers found an overall effect of zero. Our results indicate that summer Head Start is effective only for certain groups of children.

The Westinghouse study found a statistically significant negative effect when the SAT median was used as the dependent variable for the second grade summer sample. We used the SAT mean rather than the median, and we also found a significant negative effect for the second grade, both parents, summer sample; the effect was negative for both black and white children. For the second grade, summer, mothers only sample we found a positive but insignificant effect on the SAT mean for children of both races.

With regard to the full-year samples, the Westinghouse study found no significant effect of Head Start on the ITPA mean for the overall analyses of the first and second grades, but they did detect significant effects for some subgroups when the samples were stratified by region. In our reanalysis we found no significant effects on the ITPA for either race when both parents are present, but the coefficient for the first grade, mothers only present sample is significant for whites; in addition we found that although the coefficients were not significant, a positive effect of at least 1.2 points on the ITPA mean was found for blacks and whites in the full-year samples. When the MRT was used as the dependent variable, the Westinghouse study found a significant effect for Head Start in their overall analysis. In the reanalysis we found a negative effect for white children with both parents present, but a positive effect for all black children and for white children when only the mother is present. For the second grade samples we found no significant effect when the SAT is used as the dependent variable, and this finding agrees with the Westinghouse study's findings.

It is reassuring that the results of the reanalyses we have performed are consistent with the findings of the Westinghouse study; the discrepancies between the findings can be explained by the different procedures used to analyze the data. We believe that the reanalyses we have performed provide additional information that is useful in assessing the effects of Head Start.

FOOTNOTES

[1] The stratification by parents present is also desirable because total family income is likely to include more transfer payments for mother-headed families. The stratifications we have used are not exhaustive, but we have selected those stratifications that are most likely to interact with Head Start. Some of our supplementary models deal further with this topic. When the Head Start-SES interaction variables are added to the model we determine if Head Start is more or less effective for children from different backgrounds. In the section with supplementary analyses we also include a model where the data are stratified by the employment status of the mother to determine if this factor influences the effects of Head Start.

[2] For example, the mean number of children in the grade 1, summer, both parents present, black sample is 4.97 and the figure for the comparable white sample is 4.08. Mean income for the black sample is $4,725 and for the white sample it is $5,886.

[3] There is little doubt that all of the cognitive tests (which were administered in English) are culturally unfair for Mexican Americans.

[4] See Hilderth et al. (1966, p. 8).

[5] Because we have included black-Head Start and black-kindergarten interaction variables in the model, the coefficient for BLACK relates how much lower we would expect a black child who has not had kindergarten or Head Start to score than a comparable white child.

[6] It has been demonstrated that the regression framework we are employing is not the most efficient technique for determining the probabilities of group membership. Kmenta (1971, pp. 425-428) points out that the error term is heterskedastic and that a linear regression permits the fitted values of the dependent variable to be less than zero and greater than one. Kmenta suggest two modifications to solve this problem and refers the reader to the probit analysis model for a third approach [e.g., J. Tobin, "Estimation of Relationships for Limited Dependent Variables," Econometrica 26 (1958):24-36].

[7] I am grateful to Professor Harold Watts and Robert Avery for their suggestions on this topic.

[8] This result conflicts with the findings of Smith and Bissell. As we have previously noted, it is possible that Smith and Bissell applied an incorrect test for the interaction coefficients.

Chapter 6

Summary and Conclusions

1. Summary of Findings

The goal of this dissertation has been to devise a framework for analyzing the effect of Head Start on the cognitive development of disadvantaged preschool children. We have taken the view that Head Start is an input in the educational production function; we have incorporated Head Start into a linear model of that function and tested our model on the data collected for the Westinghouse Learning Corporation-Ohio University evaluation of Head Start.

Evaluations of social action programs in general, and especially compensatory education programs, are often plagued with statistical problems that lead to bias in the estimation of such programs' effect (treatment effect). In Chapter 2 we have presented several models to demonstrate the statistical problems that can occur in an evaluation of the cognitive benefits of Head Start. We have found the major problems to be that cognitive ability, prior to the commencement of the Head Start program, is measured with error, or in case of an ex post facto study is not directly available. We have found that the presence or absence of bias in the estimate of the treatment effect is dependent upon the selection procedure used to assign children to the Head Start and control groups. When the control group is selected from a population with a higher mean level of ability (as in the Campbell-Erlebacher model) or when group assignment is made on the basis of one measure of ability (with the least able children assigned to the Head Start group) and another measure is used as a control variable, we have found that the estimate of Head Start effect will be attenuated. However, certain other selection processes do not lead to bias. When there is no Head Start-ability interaction variable in the model, random selection will avoid bias. We have also demonstrated that assignment and control on the same variable--for example a pretest or measure of socioeconomic status--avoids the introduction of bias. The main lesson to be learned from the models in Chapter 2 is that to determine if a body of data will yield unbiased estimates of the treatment effect we must know not only if the Head Start and control groups differ on their pretreatment ability, but also the manner in which the children were assigned to the groups.

In Chapter 3 we have reviewed the Westinghouse study. The chapter includes discussion of the rationale for the study, the methods that were used in gathering the data, the techniques that were used to analyze the data, the statistical findings of the analysis, and the interpretations and conclusions that were made by the Westinghouse researchers. The study was carried out to assess the average impact of summer and full-year Head Start on the cognitive and affective development of children who have participated in the program. The evaluation was constrained to be ex post facto, and the control group employed was to be selected from among children in the same neighborhoods who were eligible for Head Start but did not participate in the program. The principal method of analysis was analysis of covariance with the Hollingshead Index of socioeconomic status and neighborhood dummy variables employed as covariates and the various cognitive and affective tests used as the dependent variable. Virtually all of the evidence presented in the Westinghouse report indicates that summer Head Start has been ineffective and that full-year Head Start has been only marginally effective in raising the level of cognitive development for preschool children. The Westinghouse researchers conclude that summer Head Start programs should be phased out and that full-year programs should be strengthened and continued.

Head Start has been one of the most popular programs in the war on poverty, and it is not surprising that the negative findings of the Westinghouse study have elicited many criticisms. In Chapter 4 we have reviewed the criticisms directed at the Westinghouse study and outlined the reanalysis of the data subsequently presented in Chapter 5. Some critics have argued that the questions asked by the Westinghouse researchers are not the appropriate ones for an evaluation of Head Start; they claim that by focusing on the average impact of Head Start, the study has ignored the problem of determining what leads to a successful Head Start program. We concur with the Westinghouse authors and OEO officials that for a decentralized program such as Head Start it is useful to learn the average impact as well as the most effective techniques; in addition, the planned variation study and many pilot projects have been funded to learn the efficacy of various approaches to Head Start.

Some criticism has been directed at the cognitive and background instruments utilized in the Westinghouse study, and a great deal of criticism has been made concerning the affective tests. All three of the affective measures were designed specifically for the Westinghouse study, and little is known about the validity and reliability of these instruments. Because of the weakness of the affective instruments and our lack of expertise in interpreting them, we have elected to confine our study and reanalysis to the cognitive benefits of Head Start. We felt that the *Head Start Official's Interview Questionnaire* (HSIQ) and the *Parent Interview Questionnaire* (PIQ) did not collect all of the desired information necessary for interpreting results, and in some cases the questions were poorly phrased. We therefore were unable to include all of the desired information in our reanlaysis of the data. With regard to the HSIQ, there was no question concerning the number of hours per day that the children spent at the center and questions about the curriculum and objectives of the center were phrased too broadly. The major faults in the PIQ are that the response categories for income are too broad and income is not broken down by source; the occupational codes overlap and are ambiguous. These deficiencies in the questionnaires limited our ability to include the desired control variables in the reanalysis.

The sampling procedures employed in the Westinghouse study have been criticized for the selection of both the centers to be analyzed and the individual Head Start and control children. The center selection process has been criticized because it was necessary to screen 225 centers before the final sample of 104 could be obtained; Smith and Bissell claim that the screening process employed eliminated a disproportionate number of effective centers. After reviewing the reasons for dropping centers we have concluded that no strong case can be made for the omitted centers being more or less effective than the included ones. The center selection has also been criticized because there were only 29 full-year centers in the final sample. A larger number of full-year centers would have made the estimates for full-year Head Start more efficient and would have permitted a more extensive analysis of the subgroups. (One of the OEO officials involved in the study later admitted that it was a mistake to include so few full-year centers.) The selection of the individual children in the study has been criticized because it is possible that the control children are on average more able than the Head Start children and that the analysis of covariance would undercorrect for these differences; this would then lead to an underestimate of the Head Start effect. The models presented in Chapter 2 demonstrate that differences in ability for the Head Start and control groups can lead to bias under some selection procedures, but will not in others. Thus the issue of bias cannot be resolved without knowledge of the procedure used to assign children to the Head Start and control group.

Several aspects of the statistical analysis used in the study are subject to question, and we have modified the Westinghouse researchers' analysis of covariance model in several ways for our reanalysis. The study used grouped data in their analysis, but the use of ungrouped data is preferred because it is more efficient, it permits the inclusion of variables that cannot be used when the data is grouped, and grouping appears to offer no advantages. We have also expanded the number of socioeconomic and demographic variables to be included in the analysis to include the available variables that we would expect to influence cognitive development; Head Start is only one of the determinants of cognitive development and we should include all relevant variables in our model of the educational production function. By including all relevant variables in the model we also reduce the potential bias due to imperfect measurement of pretreatment ability. A final difference between our approach to the statistical analysis and the one used in the Westinghouse study is that we have stratified the data differently. We have stratified by grade (first, second, and third), type of program (summer and full-year), and parents present (both and mother only); we have ten rather than twelve samples because there were too few observations to analyze the third grade full-year samples. In addition, we have included Head Start-ethnicity interaction variables to permit Head Start to have a different effect for various ethnic groups.

Our results for the reanalysis of the data are compatible with the Westinghouse findings, but they differ because of the different stratification used. We have found Head Start to be ineffective for white children from two-parent families, yet Head Start produced a gain of over 5 IQ points for white children from mother-headed families; these trends were observed for both the summer and full-year samples for all three grades. For the first grade samples, full-year Head Start had a positive effect for all black children, but the summer program appeared effective only for black children from two-parent families. For second and third grade samples only the second grade, full-year, mother only sample had a substantial positive effect for black children, but it was not statistically significant. Thus, full-year and summer Head Start appear to have a favorable immediate impact on black children, but the effect does not appear to be sustained; for white children Head Start appears to be effective only when the father is absent.

To determine if the Head Start and control groups differ significantly on their socioeconomic status discriminant analysis was conducted for all ten samples. Only for three of the ten samples were the Head Start and control groups significantly different at the 5 percent level, and only for the first grade, both parents, summer sample did the differences clearly favor the control group. An alternative approach to the discriminant analysis has been devised and applied to the first grade, both parents present samples. The alternative approach indicates that the Head Start children would be expected to score lower than the control children on the cognitive tests if they had not participated in Head Start; the first grade, summer, both parents Head Start children are estimated to have scored approximately 2.3 months of psycholinguistic age lower then the control group had they not participated in Head Start; for the full-year sample the estimate is 2.1 months. For most of the samples we do not have evidence of a serious selection bias with respect to observed characteristics. Whether or not there is selection bias on unobserved characteristics is a question we cannot answer. The discriminant analyses for kindergarten, however, indicate that the kindergarten children do come from significantly more advantaged backgrounds; perhaps this indicates that the kindergarten effects we have found are inflated.

Several supplementary models were tested for the first grade, both parents present samples to test some additional hypotheses. Head Start-SES interaction variables were added to the model to determine if Head Start affected children from various backgrounds in different ways; we found that the set of interaction variables had no significant effect on the explanatory power of the model. Dummy variables for neighborhood and Head Start centers were added to the model to see if performance varied across neighborhoods and if the centers varied in their effectiveness. We found that the neighborhood variables were significant but that the center dummy variables were not. There is no significant variation of the effectiveness of the different centers included in the Westinghouse sample. Another supplementary model was to regress the cognitive test score on the socioeconomic and demographic background of the children and variables for the organization, objectives, and curriculum of the Head Start center for children who had participated in Head Start. Unfortunately, important data, such as length of the program, were not available and many of the questions used to learn center characteristics were too broad to elicit differences among centers; thus, this part of the analysis produced no useful results. A final supplementary analysis was carried out to determine if Head Start is more effective for children whose mothers work; we found limited support for the hypothesis that Head Start is more effective for children from mother-headed families when the mother works rather than when she remains in the home.

2. Comparison with Other Evaluations of Preschool Programs

Although only within the past ten years has the number of preschool programs for disadvantaged children become large, there have been hundreds of evaluations of Head Start and similar programs. The reason that so many evaluations have been carried out is that compensatory education is a new field and new approaches toward structuring these programs are constantly being developed and carried out. We have not sought to review these evaluations, but have relied instead upon the summaries prepared by Marian Stearns (1971) and Lois-ellin Datta (1969) for the U.S. Department of Health, Education, and Welfare.

Datta reviews the findings of evaluations of Head Start carried out between 1965 and 1969. Addressing first the question of the short-run effect of Head Start she concludes that, "Many though not all studies of summer Head Start programs show that children's performance on general ability tests improved significantly, although the scores typically did not reach the national averages" (p. 11). For full-year programs Datta has found that, "data from the most recent [1969] studies of Full-Year programs indicate that performance tested immediately or soon after Head Start reaches the national averages on tests of general ability and learning readiness" (p. 12). In addition to problems in the design of the evaluations, Datta notes that there are at least four possible explanations of the gains achieved by Head Start children, and these can be summarized as:

> 1) the children gain in cognitive development, and the gains are attributable to the Head Start experience.
>
> 2) the gains are due to exposure to a new environment, and any experience such as kindergarten or first grade would do as well.
>
> 3) children in Head Start become familiar with the items included on the posttest and therefore score higher even though their cognitive development has not increased.

> 4) children's attitudes toward tests improve as a result of the Head Start experience and the gains on the cognitive tests are a result of the children trying to answer questions correctly after Head Start.

With regard to the long-run impact of Head Start, Datta has found that most studies indicate that the control children catch up to the Head Start children by the end of the first year of formal schooling (kindergarten or first grade). Datta and others offer several possible explanations for this outcome, but there is no generally agreed upon answer. One possibility is that the initial gains were not really gains in cognitive development but were due to exposure to the new environment, gain in familiarity with the skills tested, or a change in attitudes toward test-taking; if this is the case, then we would expect control children to catch up as soon as they come under similar influences in school. Another explanation is that the control children learn from the experience of the Head Start children; this theory is sometimes referred to as "horizontal diffusion." Other explanations involve the school and home environments. The teachers may set the standards of the classes at the lower level of the control group, and this permits them to catch up to the Head Start children. It is also possible that the Head Start children react negatively to the more structured approach in formal schooling, or that the teachers are unable to cope with the more active and inquisitive Head Start children. A final possiblity is that the Head Start children do not receive the reinforcement in the home that is necessary for continuing progress in cognitive development. Obviously the policy recommendations for Head Start depend greatly on the reasons for the decline in the effect.

The study by Stearns reviews the success of Head Start and other preschool programs for disadvantaged children conducted through 1971. Stearns's findings are similar to those of Datta; she has found that most studies of preschool programs show an immediate impact on cognitive development, but the gain is rarely as high as 10 IQ points. She also concludes that there is no critical time period between ages two and six when preschool programs are most effective. In addition, there is no conclusive evidence that the length of time that the child participates in a preschool program significantly affects the impact of the program. Stearns states that some programs are more effective than others and that, "The more a program is well-formulated, well-organized, and focused on intellectual attainment and language skills, the greater are the changes in children's intelligence test performance." We were unable to verify this finding with the Westinghouse data.

The reviews by Datta and Stearns indicate that preschool programs usually do produce small, positive immediate impacts for disadvantaged children. Our reanalysis of the Westinghouse data has led to a similar conclusion, except that we detected no gains for white children from two-parent families.

3. Policy Implications

Head Start and other preschool programs for disadvantaged children have remained popular in spite of the findings of the Westinghouse study and other evaluation efforts. However, the great majority of the evidence we have found indicates that preschool programs produce immediate gains of about 5 IQ points, and that the effects of preschool education appear to last for at most three years. We now address the questions of whether or not Head Start has met the goals established for the program and if it is worthy of continued funding.

When Head Start was originally planned, the expectations were quite high. Stearns (1971, p. 2) summarizes the underlying philosophy as follows:

> The idea [of Head Start] was to nip failure in the bud with a multi-faceted boost in a single generation. The child who thus got an equal start, physically, emotionally, and intellectually, as he entered school would not be subject to failure, would not fall further and further behind his peers, would not drop out and become unemployable, and thus would not bring up another generation in poverty.

It is clear that Head Start and other preschool programs have not come close to meeting these lofty goals, but this does not imply that preschool education should be dismissed as a failure. Head Start has produced some immediate gains in the cognitive development of disadvantaged children, and we must determine if these benefits are worth the expenditures of the program. Considering the great popularity of Head Start, if the cognitive gains are not viewed as being great enough to justify the costs, efforts should be made to improve the cognitive components of the program.

It is apparent that unlike many other preschool programs, Head Start did not emphasize the cognitive aspects of the program. Stearns (1971, p. 117) claims that:

> The Head Start goals . . . did not include school preparation, much less emphasize it. And it is clear that the majority of Head Start centers did not take academic achievement as a primary goal. It was de-emphasized in all suggested curricula for Head Start programs issued in the Federal guidelines.

Although we have found in our review of the evaluation literature that no one particular curriculum is superior to any other, the consensus is that cognitively-oriented programs do produce larger cognitive gains than other types of preschool programs. Two of the leading developers of preschool programs, David P. Weikart and Carl Bereiter, have reached this same conclusion in the Stanley volume (1972). Thus, one policy recommendation is that if enhancing the cognitive development of children is to remain a goal of Head Start more emphasis should be placed on the use of cognitively-oriented curricula. To retain the decentralized administration of Head Start the Office of Child Development could alter the guidelines for programs and make many curricula available to the centers. In this way Head Start could be made as effective as other preschool programs while retaining its unique combination of local autonomy and federal guidance.

The conclusion of the Westinghouse study that summer Head Start programs should be phased out and replaced by full-year programs does not appear warranted by our reanlaysis. Although we discovered some instances where full-year programs appear to be more effective, these differences are not great enough to merit the large increase in funding required. Perhaps as the Head Start curricula become more effective the need for more full-year programs will be more clearly demonstrated.

No matter how large an immediate impact is produced by preschool programs, they cannot really be considered sucessful if the gains are depleted within a short period of time. The policy recommendations for sustaining initial gains will depend upon the reasons for the subsequent loss. If the initial gains are simply an artifact due to the initial change of environment or changes in the children's attitudes toward taking tests, then efforts must be made to develop programs that produce real gains or preschool programs should abandon the attempt to affect cognitive development and concentrate on other aspects of development. Stearns suggests that programs where intervention begins at even younger ages should be tried as well as programs which focus on the children when they enter first grade. Some researchers, such as Raymond S. Moore and Dennis R. Moore (1972) claim that children are not mentally and physically developed enough at age five to begin learning to read and count; they advocate that programs stressing cognitive development be avoided for children under six. If the gains from preschool programs fade because the public schools do not reinforce the preschool experience (or negatively reinforce it) then other actions are called for. Stearns (1971, p. 145) suggests three solutions to this problem:

> 1) change the goals, methods, and/or content of public school programs, 2) make preschool programs more compatible with existing primary programs, 3) make both the preschool and primary programs plan a sequence of experiences under compatible philosophies.

A final possible explanation suggested by Stearns for the decay of cognitive gains is that the home environment does not suitably reinforce the preschool experience. If this is the problem then attempts should be made to modify the home environment, to remove the children from the home environment for a longer period of time, or to modify preschool programs so that the existing home environment can be utilized to retain the effects of the program. Thus, further research is needed to determine why preschool programs do not produce more permanent gains.

Preschool education remains an area where a great deal of additional research is needed. We do not agree with the statement by Jensen (1969) that "Compensatory education has been tried and it apparently has failed." His conclusion is premature, as the field is too new to dismiss at this time. Preschool education can produce some gains and research should continue until we learn how to sustain and enhance these gains. At the same time we must realize the limitations of preschool education. As Jencks et al. (1972) have correctly claimed, education will not eliminate inequality in our society. It is likely, however, that Head Start and other preschool education programs can play an important part in the education of disadvantaged children.

Appendix

Additional Empirical Results

Because of the large number of samples we have considered and the great number of cognitive measures available from the Westinghouse data, we have included much of our analysis in this appendix rather than in the text. This has been done mainly to spare the reader from the task of wading through an additional 59 tables. A brief summary of the Appendix appears in Chapter 5. There are four groups of tables in the Appendix and a description of each group appears below.

1. Analyses of the Subtests for the First Grade, Summer Both Parents Sample (Tables A-1 through A-19)

In the Westinghouse study the analyses of covariance were performed for every sample using each of the subtests for the cognitive measures as well as the total test scores as the dependent variables. To keep this dissertation at a reasonable length, we have reported only regressions using the mean of the cognitive tests in the text. Although Hilderth et al. (1966) and Kirk et al. (1968) discuss the total scores of the MRT and ITPA as being the sum of the subtest scores, we have used the means in our analyses to make the scale comparable for the total score and the subtest scores; this linear transformation of the dependent variable only affects the scale of the regression coefficients and has no effect on the statistical significance of the regression coefficients or on the coefficient of determination. The only other modifications we have made in the use of the test scores is that we have employed the mean of the nonzero subtest scores and we have skipped all observations where three or more of the subtests on any test are equal to zero. This procedure has been used because zero scores indicate that either the child has not taken the subtest, or he is so retarded that he is not of interest for our evaluation.

Table A-1 gives the correlations among the ITPA and MRT subtests and means for the first grade, summer, both parents sample. All of the correlations are positive, and the coefficients range in size from .125 to .869. The correlation between the ITPA mean and the MRT mean is .608. Hilderth et al. (1966, p. 12) present the correlation coefficients for the MRT subtests that were found on a sample of 12,225 children; the correlations in Table A-1 are fairly close to those reported by Hilderth et al. The MRT manual's claim that, "none [of the correlations] is so large as to suggest that any two of the tests are measuring identical or nearly identical functions" is subject to question. Low correlation coefficients can be an indication that the individual tests are highly fallible, and high correlations could be indicative of a high correlation of the various skills measured by the tests.

The regressions of the subtest scores and the mean scores on the set of socioeconomic and demographic characteristics are contained in Tables A-2 through A-19. As we would expect when the dependent variables are highly correlated measures of similar skills, the coefficient for Head Start when the mean of the subtests is used as the dependent variable is approximately equal to the mean of the coefficients for Head Start when the subtests are used; for example, the coefficient for Head Start for whites when the ITPA mean is used as the dependent variable is -.361, and the mean of the coefficients for the subtest effects is -.365. As we have found for the regressions discussed in chapter 5, the coefficients for many of the socioeconomic and demographic variables are somewhat unstable, and the coefficient for a variable may be positive for one subtest and negative for another. The regressions in Tables A-2 and A-13 are also found in Tables 5.4 and 5.5 in the text and are also reported in Tables A-24 and A-25; this repetition has been done to facilitate comparison between various groups of regression equations.

2. Regressions of Mean Test Scores on Individual Characteristics for All Samples (Tables A-20 through A-39)

The next group of tables contains the regression equations with the ITPA mean and the achievement test mean (MRT, SAT2, or SAT3 depending on the grade level) used as the dependent variables for the ten samples. As we have mentioned in the text, the data were stratified by grade (first, second, and third), type of program (summer and full-year), and parents present (both or mother only); there are ten rather than twelve samples because there were too few third grade, full-year children to analyze. Tables A-20, A-21, A-24, and A-25 are also contained in the text (Tables 5.4 and 5.5) and need not be discussed again; the Head Start and kindergarten coefficients of the regressions using the ITPA mean as the dependent variable are listed in Table 5.6 of the text.

Regressions for the samples where only the mother is present are not reported in the text because there were small sample sizes and the coefficients did not differ much from those found in the samples where both parents are present. For these samples the variables for the education and occupation of the father have been omitted from the equation as they are not relevant. Variables for the marital status of the mother have been added to determine if the reason for the absence of the father affects cognitive development. The reference group for the marital status variables is divorced mothers, and the other categories are women who are separated from their husbands (SEPAR), widows (WIDOW), and women who were never married (NEVMAR). Table A-22 gives the regression equation for the first grade, full-year, mother only sample when the ITPA mean is used as the dependent variable. None of the marital status coefficients are statistically significant, and only the coefficient for widows is positive; in Table A-23 where the MRT mean is used as the dependent variable, all of the marital status coefficients are positive, but insignificant. The coefficients for the marital status variables are insignificant for all 20 regressions reported for samples where only the mother is present, and the signs of the coefficients vary from sample to sample.

Two versions of the Stanford Achievement Test (SAT2 and SAT3) were used as cognitive measures for the second and third grade samples in addition to the ITPA, which was administered to all children. We have only reported the Head Start and kindergarten coefficients for the regressions using the ITPA in the text because only the ITPA can be used to make intergrade comparisons, and it is easier to interpret the coefficients when the ITPA is used. It appears that both we and the Westinghouse researchers have scaled the SAT incorrectly. Due to some confusion about the coding of the SAT scores, the SAT scores used in our analyses are multiplied by a factor of ten times the true score. The SAT scores reported in the Westinghouse study are one-tenth of the true scores; although they do not discuss this, it is apparent from examining the reported scores and the average scores in the SAT manuals. These differences in scale do not affect the findings other than changing the regression coefficients by a factor of ten.

The SAT manuals, Kelly et al. (1966) and (1965), recommend the use of the battery median as a measure of total achievement; because both the SAT2 and SAT3 have an even number of subtests, the total score is actually the mean of the two middle-ranked subtests. We have used the battery mean in our reanalysis because the mean better reflects performance on all subtests. The SAT manuals offer only a percentile chart for interpreting the scores. For the way in which we have scaled the SAT scores a Head Start effect of 10.0 SAT points is approximately equal to a gain of 7.5 percentiles for children at the level of those in the Westinghouse samples for the second grade, and a coefficient of 10.0 for the third grade samples is approximately equal to a gain of 4.0 percentiles. As mentioned in the text, the only significant coefficient for Head Start that we found is a negative one for the second grade, summer, both parents sample (see Table A-33); the Westinghouse researchers also found a significant, negative coefficient for a comparable sample. When the ITPA mean was used as the dependent variable for this sample, a negative coefficient for Head Start was found, but it was not significant (see Table A-32).

3. Head Start Discriminant Analyses (Tables A-40 through A-49)

As we have mentioned in the text, discriminant analysis has been used to determine if the Head Start and control children differ significantly in their socioeconomic and demographic backgrounds. The discriminant analyses present evidence concerning whether the regression coefficients are biased, but we can make no definite statement concerning the bias issue based on the discriminant analyses. This is because some important background variables (such as parental attitudes) were not available for the analysis, and differences in background are not a sufficient condition for bias. For only two of the ten samples can the hypothesis that the Head Start and control groups have the same means on the background variables be rejected at the 5 percent level. These two samples are the first grade, full-year, mother only sample (Table A-41), and the first grade, summer, both parents sample (Table A-42). In both cases it appears that the control group comes from a more advantaged background. Quite surprisingly, a positive effect for Head Start was found for the full-year sample (statistically significant for white children), and a positive effect was found for black children in the summer sample (see Tables A-22, A-23, A-24, and A-25). For the other eight samples there is no clear evidence of control group superiority, and this reduces the likelihood of our understating the effects of Head Start in our regressions.

4. Kindergarten Discriminant Analyses (Tables A-50 through A-59)

Because we found such consistently high coefficients for kindergarten and because no attempt was made in the Westinghouse study to match the children who attended kindergarten with those who did not, we have also conducted discriminant analyses for all ten samples using the dummy variable for kindergarten attendance as the dependent variable. The results of these analyses differ sharply from the Head Start discriminant analyses. The F-statistics are significant for all ten samples, and the coefficients clearly indicate that the children who attended kindergarten came from more advantaged bakcgrounds. The coefficient for income is positive for all ten samples, and it is statistically significant at the 5 percent level for eight of the samples.

There is a great deal of information that we have not included in the text or this appendix because it is of interest only to a small number of readers. The means and standard deviations of all variables for all samples are available upon request from the author.

Table A-1

Correlation Matrix of All Test Scores for Grade 1, Summer, Both Parents Sample

	AUDREC	VISREC	VISSEQ	AUDASS	AUDSEQ	VISASS	VISCLO	VEREXP	GRACLO
AUDREC	1.000	.387	.183	.581	.165	.371	.292	.384	.507
VISREC		1.000	.238	.451	.125	.410	.373	.316	.410
VISSEQ			1.000	.251	.138	.273	.290	.207	242
AUDASS				1.000	.391	.452	.380	.491	.661
AUDSEQ					1.000	.153	.171	.210	.237
VISASS						1.000	.403	.302	.371
VISCLO							1.000	.374	.337
VEREXP								1.000	.479
GRACLO									1.000

Table A-1 (cont.)

	AUDREC	VISREC	VISSEQ	AUDASS	AUDSEQ	VISASS	VISCLO	VEREXP	GRACLO
MANEXP	.247	.353	.176	.323	.121	.308	.380	.453	.309
ITPAMN	.669	.640	.440	.792	.472	.621	.631	.687	.716
WORD	.371	.330	.106	.407	.156	.295	.148	.246	.362
LIST	.419	.366	.179	.479	.189	.357	.263	.282	.458
MATCH	.346	.395	.229	.471	.224	.372	.329	.231	.360
ALPHA	.316	.257	.227	.391	.242	.263	.158	.193	.313
NUMB	.382	.367	.283	.497	.268	.369	.273	.325	.437
COPY	.300	.293	.291	.391	.210	.356	.260	.163	.320
MRTMN	.461	.432	.290	.574	.294	.434	.309	.316	.491

Table A-1 (cont.)

	MANEXP	ITPAMN	WORD	LIST	MATCH	ALPHA	NUMB	COPY	MRTMN
MANEXP	1.000	.580	.125	.203	.131	.123	.203	.136	.198
ITPAMN		1.000	.413	.512	.492	.397	.546	.430	.608
WORD			1.000	.577	.382	.330	.489	.308	.659
LIST				1.000	.511	.352	.585	.424	.724
MATCH					1.000	.448	.600	.521	.756
ALPHA						1.000	.574	.483	.740
NUMB							1.000	.571	.869
COPY								1.000	.715
MRTMN									1.000

Table A-1 (cont.)

Abbreviations used in the correlation matrix.

AUDREC = Auditory Reception ITPA Subtest

VISREC = Visual Reception ITPA Subtest

VISSEQ = Visual Sequential Memory ITPA Subtest

AUDASS = Auditory Association ITPA Subtest

AUDSEQ = Auditory Sequential Memory ITPA Subtest

VISASS = Visual Association ITPA Subtest

VISCLO = Visual Closure ITPA Subtest

VEREXP = Verbal Expression ITPA Subtest

GRACLO = Grammatic Closure ITPA Subtest

MANEXP = Manual Expression ITPA Subtest

ITPAMN = ITPA Mean

WORD = Word Meaning MRT Subtest

LIST = Listening MRT Subtest

MATCH = Matching MRT Subtest

ALPHA = Alphabet MRT Subtest

NUMB = Numbers MRT Subtest

COPY = Copying MRT Subtest

MRTMN = MRT Mean

Table A-2

Effects of Individual Characteristics on ITPA Mean

Independent Variable	Coefficient	T-Ratio
CONSTANT	15.66257562	8.42085
CHILD	-.13433372	-1.89476
INCOME	-.00001000	-.17300
AGE	.76607500	2.79095
MHSG	-.13460522	-.34511
MSOMC	-.50832257	-.85904
MCG	-.77554412	-1.14506
MGS	-1.62245730	-1.97430
MOCLER	2.49301492	1.97716
MOCKIL	1.87575447	1.50934
MOCSMI	.15878783	.14829
MOUNSK	.50540495	.51810
FEMALE	-.13315325	-.43300
RURAL	.24025747	.70124
FHSG	.34161130	.63055
FSOMC	-.20333331	-.00541
FCG	-.47537477	-.88991
FGS	-.90460000	-1.45284
FACLER	-.93128783	-1.07955
FACKIL	-1.00610727	-1.63376
FACSMI	-1.15820160	-1.89547
FAUNSK	-1.73561152	-2.57357
BLACK	-2.10324027	-3.11208
MEXAM	-1.12157257	-1.68794
KIND	.95135275	2.50445
BLKIND	-.59326165	-.89494
SUMMER	-.36060374	-1.00394
BLKHS	1.99702473	3.12482
MEXHS	1.57234515	1.77683

$R^2 = .1860$

Note: The sample (623) in this table and for results reported through Table A-19 includes those from the first grade with both parents present in the summer Head Start program.

Table A-3

Effects of Individual Characteristics on ITPA Auditory Reception Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	19.91403070	5.21107
CHILD	-.27022237	-1.68925
INCOME	.00006021	.54204
AGE	.05159366	1.80426
MHSG	-.72083374	-.66705
MSOHS	-1.05602490	-.94381
M78	-1.29325465	-1.03921
MSE	-2.49315627	-1.57881
MOCLER	3.46597747	1.47182
MOCSKIL	1.73033491	.74545
MOCSMI	.47147545	.22981
MOCUNSK	1.23335645	.68759
FEMALE	-.16255211	-.30495
RURAL	.75917333	.54796
FHSG	.35532005	.26894
FSOHS	.07125422	.07120
F78	-.15830510	-.16456
FSE	-.71524265	-.26679
FACLER	-3.74357337	-2.28341
FASKIL	-3.04260515	-2.57105
FASEMI	-3.08556059	-2.60521
FAUNSK	-3.73370433	-3.09743
BLACK	-4.33457569	-3.33816
MEXAM	-4.45385070	-3.49631
KIND	1.70880750	1.98052
BLKIND	-1.00540421	-.78925
SUMMER	-.93177009	-.99766
BLKHS	2.71527461	2.22218
MEXHS	2.30465494	1.39071

$R^2 = .1604$

Table A-4

Effects of Individual Characteristics on ITPA Visual Reception Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	15.[illegible]	5.40319
CHILD	-.[illegible]	-1.88956
INCOME	.[illegible]	.34999
AGE	.[illegible]	1.53474
MHSG	-1.[illegible]	-1.23450
MSOHS	-.[illegible]	-.50629
MTR	-.[illegible]	-.72522
MCC	-1.[illegible]	-1.20477
MOCLER	1.[illegible]	.88574
MOCKIL	-1.[illegible]	-.61293
MOCCMT	-1.[illegible]	-.71017
MOUNCK	-.[illegible]	-.29077
FEMALE	-.[illegible]	-1.61319
RURAL	.[illegible]	.61796
FHSG	.[illegible]	.18387
FSOHS	-.[illegible]	-.40126
FTR	-1.[illegible]	-1.23747
FCC	-.[illegible]	-.51832
FACLER	-.[illegible]	-.27920
FACKIL	-.[illegible]	-.90894
FACCMT	-1.[illegible]	-1.06033
FAUNCK	-1.[illegible]	-1.76616
BLACK	-1.[illegible]	-1.07049
MEXAM	-.[illegible]	-.59019
KIND	1.[illegible]	1.99090
BLKIND	-.[illegible]	-.75668
SUMMER	.[illegible]	.00868
BLKHS	1.[illegible]	1.20891
MEXHS	1.[illegible]	1.10134

R^2 = .0881

Table A-5

Effects of Individual Characteristics on ITPA Visual Sequential Memory Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	15.77690005	5.89921
CHILD	.07635297	.38556
INCOME	-.00001395	-.17914
AGE	.33004202	1.05652
MHSG	-.11131206	-.14747
MSOHS	-.55308050	-.70138
M78	-.14041555	-.15192
M56	-.33340637	-.35156
MOCLER	-.04348015	-.02632
MOCKIL	.07675277	.02170
MOCSMT	-.60674771	-.48382
MOUNSK	-.60255701	-.50388
FEMALE	.11337344	.30468
RURAL	-.25870715	-.58105
FHSG	.77325826	.50979
FSOHS	.67251432	.95798
F78	-.17340467	-.24081
F56	-.42038649	-.51135
FACLER	1.46215740	1.27117
FACKIL	.95323309	1.14980
FACSMT	.37522087	.45157
FAUNSK	.53457157	.62391
BLACK	-1.01279375	-1.33130
MEXAM	2.23006815	2.54543
KIND	-.07460991	-.07459
BLKIND	-.27756652	-.31057
SUMMER	.15602003	.34281
BLKHS	.54415241	.63471
MEXHS	.10855264	.09337

R^2 = .0587

Table A-6

Effects of Individual Characteristics on ITPA Auditory Association Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	12.79170854	4.15477
CHILD	-.07138846	-.28468
INCOME	.00003678	.40927
AGE	1.65557102	2.47599
MHSG	.07019482	.08078
MSOHS	-.61127571	-.67217
M79	-1.46315657	-1.46707
MC6	-1.35547887	-1.45221
MOCLER	2.31572961	1.74029
MOSKIL	1.54779665	.80698
MOSEMI	.41200252	.24923
MOUNSK	.41758027	.27534
FEMALE	.41226654	.95514
RURAL	.69951251	1.31716
FHSG	.21546271	.28019
FSOHS	.30361158	.49232
F79	-.42249657	-.50976
FC6	-1.32323094	-1.43970
FACLER	-.73162303	-.59677
FASKIL	-1.22757479	-1.29260
FASEMI	-.93055527	-.98047
FAUNSK	-1.52900080	-1.54743
BLACK	-2.92732230	-2.69175
MEXAM	-2.44561910	-2.36687
KIND	2.35711734	4.42382
BLKIND	-2.33501370	-2.26640
SUMMER	-.25870326	-.46322
BLKHS	2.45370110	2.48194
MEXHS	.27595574	.20581

$R^2 = .2020$

Table A-7

Effects of Individual Characteristics on ITPA Auditory Sequential Memory Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	18.34628702	4.30811
CHILD	.14409323	.94203
INCOME	.00002779	.22217
AGE	.23929477	.50557
MHSG	1.03022710	.89875
MSOHS	-.71768794	-.56987
M78	-.77201738	-.52042
MGS	-1.37599201	-1.05940
MOCLER	2.61824347	.99056
MOSKIL	3.32659126	1.24091
MOSEMI	1.89377046	.82242
MOUNSK	2.05365048	.97608
FEMALE	1.87758527	3.07030
RURAL	1.06427029	1.44009
FHSG	1.03607257	.97118
FSOHS	.64004029	.56982
F78	.97296544	.84455
FGS	-1.60067377	-1.19179
FACLER	-1.34197541	-.56622
FASKIL	-1.93800277	-1.45904
FASEMI	-2.49734011	-1.87525
FAUNSK	-3.24607302	-2.36808
BLACK	1.90458291	1.23974
MEXAM	-.50658529	-.42334
KIND	.35591799	1.29099
BLKIND	-.77277391	-.54047
SUMMER	-.30585749	-1.16909
BLKHS	3.30239277	2.40795
MEXHS	.77849710	.41948

$R^2 = .1320$

Table A-8

Effects of Individual Characteristics on ITPA Visual Association Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	17.70195890	8.61953
CHILD	-.33471070	-3.90902
INCOME	-.00002317	-.28770
AGE	.85525460	2.33227
MHSG	.24331314	.32007
MSCHS	.11837825	.17602
MZ9	-.35737570	-.41104
MDE	-1.63231319	-1.46996
MOCLER	2.54132495	1.53449
MOSKIL	-.30343957	-.19062
MOSEMI	-.20227792	-.14018
MOUNSK	-.17135714	-.10001
FEMALE	-.45338547	-1.25039
RURAL	.43798280	1.05758
FHSG	.52307192	.85709
FSCHS	-.15202816	-.21595
FZ9	.03679977	.05097
FDE	-.23358930	-.28346
FACLER	-1.04102067	-.90337
FASKIL	-1.57487227	-1.84780
FASEMI	-1.13002457	-1.42955
FAUNSK	-1.20397023	-1.40154
BLACK	-2.61591487	-2.86403
MEXAM	1.24009341	1.38169
KING	1.37460300	2.86910
BLKING	-.65309012	-.72053
SUMMER	-.25773137	-.52995
BLKHS	1.93516254	2.13515
MEXHS	.54146714	.55029

R^2 = .1593

Table A-9

Effects of Individual Characteristics on ITPA Visual Closure Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	11.79408778	3.53279
CHILD	-.14169204	-1.18417
INCOME	.00000760	.65216
AGE	1.71750533	2.84549
MHSG	.25505535	.26985
MSOHS	.51860358	.62693
M79	.21675411	.19977
MCS	-.97710744	-.70473
MCCLER	-.41955299	-.20749
MCSKIL	2.45447454	1.17048
MCSEMI	-.59299741	-.32368
MCUNSK	.03454165	.02099
FEMALE	.00000034	.02119
RURAL	-.64000574	-1.10713
FHSG	-.84914794	-1.01754
FSOHS	-.45671219	-.51990
F79	-1.28475773	-1.42563
FCS	-1.73402009	-1.65048
FACLER	.12943363	.08992
FASKIL	-.17835257	-.17214
FASEMI	-.69620319	-.65998
FAUNSK	-1.74501355	-1.62741
BLACK	-1.07975261	-.94709
MEXAM	2.47957499	2.21199
KIND	1.14907845	1.97990
BLKIND	-.03629400	-.07714
SUMMER	.27412403	.45229
BLKHS	.84577119	.78928
MEXHS	.60385759	.41501

R^2 = .1010

Table A-10

Effects of Individual Characteristics on ITPA Verbal Expression Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	17.3057645	4.42251
CHILD	-.1199114	-.83272
INCOME	-.3001195	-1.00797
AGE	.4743457	.85046
MHSG	-.7051707	-.61523
MSCHS	-.5308330	-.53005
M79	-1.5244150	-1.24117
MCE	-2.2900347	-1.24398
MOCLER	5.2155250	2.73774
MOCKIL	7.7003250	3.04493
MOCEMT	2.0721680	.95397
MOUNCK	2.1841451	1.10044
FEMALE	-.5917715	-1.04741
RURAL	.4255771	.60296
FHSG	.5420344	.53959
FSCHS	-.1829543	-.17257
F79	-1.4175591	-1.30436
FCE	-1.6567319	-1.30763
FACLER	-1.9491370	-1.12277
FACKIL	.0613596	.04897
FACEMT	-1.7500260	-1.09229
FAUNCK	-.8727251	-.57496
BLACK	-3.7472497	-2.43454
MEXAM	-2.3390265	-2.10007
KIND	.2375755	.33949
BLKIND	.0659527	.04890
SUMMER	-.5674055	-.91311
BLKHS	3.2341024	2.34509
MEXHS	1.4121575	.80978

R^2 = .1229

Table A-11

Effects of Individual Characteristics on ITPA Grammatic Closure Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	13.39067719	4.54004
CHILD	-.14714151	-1.50428
INCOME	-.00001195	-.15001
AGE	1.07316340	2.83481
MHSG	-.52535630	-.67790
MSCHS	-.75605737	-.93714
MZR	-1.57289459	-1.77299
MGE	-2.15402357	-1.90038
MOCLER	2.86699407	1.69616
MOSKIL	4.52305304	2.63844
MOSEMI	1.15136164	.78906
MOUNSK	1.70433367	1.26673
FEMALE	.07844519	.20496
RURAL	.23750462	.50257
FHSG	.60962491	.89360
FSCHS	-.60627436	-.84405
FZR	-.50701740	-.68821
FGE	-.53190913	-.79383
FACLER	-1.59845787	-1.44328
FASKIL	-2.22455117	-2.61876
FASEMI	-1.60914807	-1.89285
FAUNSK	-2.33456409	-2.66327
BLACK	-4.57150056	-4.96943
MEXAM	-4.85025451	-5.30747
KIND	-.01137135	-.02395
BLKIND	-.22698579	-.24825
SUMMER	-.77377289	-1.56169
BLKHS	2.18051208	2.48619
MEXHS	2.45536418	2.06423

$R^2 = .2518$

Table A-12

Effects of Individual Characteristics on ITPA Manual Expression Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	22.5489276	7.16708
CHILD	-.1835806	-1.62807
INCOME	-.0001013	-1.10228
AGE	.4979421	1.14100
MHSG	.0657534	.07465
MSOHS	-.4889166	-.52570
MZ8	.2053065	.20074
MCG	-.3611579	-.65906
MOCLER	1.0900749	.55943
MOSKIL	-1.3218237	-.66887
MOSEMI	-1.3190094	-1.13059
MOUNSK	-1.7845563	-1.15057
FEMALE	-1.5517983	-3.51716
RURAL	-.1707722	-.31347
FHSG	.5504114	.82703
FSOHS	-.0265681	-.02424
FZ8	-.9307113	-.97914
FCG	-.5379288	-.59381
FACLER	-.1647081	-.12148
FASKIL	-.1642055	-.16769
FASEMI	.3176529	.32413
FAUNSK	-.2673387	-.26456
BLACK	-1.3681190	-1.27336
MEXAM	-1.0902659	-1.03201
KIND	.5139674	.93923
BLKIND	-.2922072	-.27722
SUMMER	-.4407151	-.77160
BLKHS	1.8513237	1.78171
MEXHS	2.7851409	2.03198

$R^2 = .0767$

Table A-13

Effects of Individual Characteristics on MRT Mean

Independent Variable	Coefficient	T-Ratio
CONSTANT	5.43711155	3.81515
CHILD	-.57343987	-1.42454
INCOME	.00002959	.70658
AGE	.67352709	3.37610
MHSG	-.19527197	-.47768
MSOHS	-.62507717	-1.47259
M79	-.92160294	-1.97120
MSS	-1.54618075	-2.59855
MOCLER	1.73264345	2.01251
MOSKIL	.36343300	.95581
MOSEMI	.77445200	.99805
MOUNSK	.97245169	1.23049
FEMALE	.41937747	2.07928
RURAL	.22370087	.89825
FHSG	.59065704	1.64293
FSOHS	.28360260	.74923
F79	.44294462	1.14065
FSS	-.10090451	-.22272
FACLER	-.54547014	-.88118
FASKIL	-.79501304	-1.77600
FASEMI	-1.09755597	-2.45013
FAUNSK	-1.22614555	-2.65433
BLACK	-1.55154168	-3.15895
MEXAM	-.72449779	-1.50015
KIND	.77001020	3.07813
BLKIND	-.57810556	-1.19977
SUMMER	-.24472551	-.93727
BLKHS	.95439412	2.06494
MEXHS	.79619930	1.27004

R^2 = .2088

Table A-14

Effects of Individual Characteristics on MRT Word Meaning Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	7.504174972	4.43252
CHILD	-.179815799	-2.92453
INCOME	-.000004117	-.08206
AGE	.475554495	2.00264
MHSG	-.271329577	-.55645
MSOHS	-.539220644	-1.06132
M79	-.953633516	-1.71001
MCE	-.971693448	-1.36391
MOCLER	1.578075135	1.48526
MOSKIL	.071291431	.06615
MOSEMI	.092957795	.10043
MOUNSK	.304023037	.35948
FEMALE	-.302809117	-1.25966
RURAL	.545745940	1.83717
FHSG	.398845411	.93008
FSOHS	.185900321	.41173
F79	.568366374	1.22733
FCE	-.599404992	-1.11025
FACLER	-1.166823352	-1.57738
FASKIL	-1.403991145	-2.62943
FASEMI	-1.562379127	-3.11099
FAUNSK	-1.149477621	-2.08614
BLACK	-.977153517	-1.49722
MEXAM	-1.056851042	-1.85200
KIND	.540297492	2.14585
BLKIND	-1.591455655	-2.76995
SUMMER	-.278441993	-.89403
BLKHS	.979142263	1.77605
MEXHS	.355374417	.47663

$R^2 = .1735$

Table A-15

Effects of Individual Characteristics on MRT Listening Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	—	—
CHILD	7.69318120	4.88671
INCOME	-.08291274	-1.46947
AGE	.00007396	.73520
MHSC	.58762581	2.69049
MSOHS	-.59059158	-1.54335
MZ9	-.71933103	-1.54570
MOS	-1.16889300	-2.28387
MOCLER	-1.07160587	-1.63897
MOSKIL	.81614319	.83705
MOSEMI	.74746756	.75184
MOUNSK	.13418579	.15798
FEMALE	.03143430	.10493
RURAL	.07879008	.35694
FHSC	-.02302986	-.30458
FSOHS	.33437175	.57016
FZ9	-.00021933	-.00053
FOS	.23564379	.55451
FACLER	-.36952544	-.74586
FASKIL	.47383142	.69802
FASEMI	-.22655172	-.46236
FAUNSK	-.44812012	-.91396
BLACK	-.67245520	-1.32990
MEXAM	-1.78243174	-3.31540
KIND	-1.79734260	-3.40000
BLKIND	.45007463	1.64367
SUMMER	-.59267305	-1.12370
BLKHS	-.33074651	-1.15724
MEXHS	.91971455	1.81792
——	1.57497255	2.29527

R^2 = .1668

Table A-16

Effects of Individual Characteristics on MRT Matching Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	2.79317136	1.43089
CHILD	-.03719704	-.53167
INCOME	.00002094	.36579
AGE	.51303068	2.26421
MHSG	-.80493510	-1.45059
MSCHS	-.30082764	-1.56111
M78	-1.33257475	-2.09998
MDS	-2.39137265	-2.94972
MOCLER	3.27981204	2.67990
MOSKIL	3.08869562	2.51903
MOSEMI	2.97966235	2.73425
MOUNSK	2.71596295	2.82216
FEMALE	.73165754	2.67277
RURAL	-.14535004	-.43001
FHSG	1.05340090	2.15983
FSCHS	.40972432	.79762
F78	.93311576	1.77093
FDS	.25291202	.42798
FACLER	.22270069	.26458
FASKIL	-.26268735	-.43236
FASEMI	-.55859429	-1.08329
FAUNSK	-1.19158000	-1.90053
BLACK	-1.48461458	-2.22707
MEXAM	.28200751	.43023
KIND	.95425597	2.81232
BLKIND	-.57965650	-.82365
SUMMER	-.20389563	-.57532
BLKHS	.02433483	.03879
MEXHS	.42396715	.49833

R^2 = .1668

Table A-17

Effects of Individual Characteristics on MRT Alphabet Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	—	
CHILD	6.05264776	2.37842
INCOME	-.19425182	-2.12990
AGE	.00007785	1.04717
MHSG	.77152226	2.18567
MSGHS	.05957270	.08237
M79	-.46818307	-.61173
MSS	-1.01469398	-1.22657
MOCLER	-1.65060221	-1.56175
MOSKIL	1.13242899	.75658
MOSEMI	.27634572	.17319
MOUNSK	-.02201783	-.01604
FEMALE	.19123425	.15304
RURAL	1.17192317	3.28357
FHSG	.73870845	1.67658
FSGHS	.55567138	.87353
F79	.44806017	.65704
FSS	.30157267	.52635
FACLER	-.24646447	-.30775
FASKIL	-1.13362562	-1.03311
FASEMI	-1.00358078	-1.26705
FAUNSK	-1.49685725	-1.88833
BLACK	-1.04127085	-1.27391
MEXAM	-1.47915673	-1.70204
KIND	-1.27813924	-1.44900
BLKIND	1.07832320	2.34583
SUMMER	.32442741	.38053
BLKHS	-.21970270	-.47379
MEXHS	1.33587429	1.63351
---	2.35401089	2.12741

$R^2 = .1172$

Table A-18

Effects of Individual Characteristics on MRT Numbers Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	6.16683697	2.46092
CHILD	.04701421	.52346
INCOME	.00004598	.62810
AGE	.98814463	2.84277
MHSG	.33086954	.54877
MSOHS	-.58343619	-.78760
MZ9	-.47378252	-.58161
MOE	-1.96526755	-1.88832
MOCLER	2.49892717	1.60361
MOSKIL	1.08919006	.69196
MOSEMI	.85702304	.63389
MOUNSK	1.14840729	.92959
FEMALE	.39109399	1.11289
RURAL	-.07606259	-.17529
FHSG	.92471835	1.47623
FSOHS	.44973662	.68190
FZ9	.44249521	.65414
FOE	-.24569819	-.31155
FACLER	-1.11174629	-1.02898
FASKIL	-1.57597791	-2.02053
FASEMI	-1.70255332	-2.18115
FAUNSK	-2.21232107	-2.74875
BLACK	-2.99398316	-3.49857
MEXAM	-1.35171325	-1.48754
KIND	.54202751	1.47300
BLKIND	-.45104254	-.53724
SUMMER	-.45967920	-1.01042
BLKHS	1.54469904	2.04233
MEXHS	.30896992	.83225

$R^2 = .1621$

Table A-19

Effects of Individual Characteristics on MRT Copying Subtest

Independent Variable	Coefficient	T-Ratio
CONSTANT	1.70167548	.88257
CHILD	.03208347	.46434
INCOME	.00002712	.41045
AGE	.75230405	2.81580
MHSG	.12875791	.23580
MSCHS	-.37863666	-.66440
M72	-.93350510	-1.48956
M66	-1.77939962	-2.22315
MOCLER	1.99646474	1.67209
MOCKIL	.13303027	.10986
MOCEMT	.57879171	.61414
MOUNCK	.83649211	.88513
FEMALE	.57824955	2.50973
RURAL	.05652942	.15158
FHSG	.40825237	.84719
FSCHS	.00011595	.00023
F72	-.20539337	-.39593
F66	.23398300	.48459
FACLER	-1.25315740	-1.51954
FACKIL	-.97073307	-1.39881
FACEMT	-.97535575	-1.62423
FAUNCK	-1.59637015	-2.57908
BLACK	-1.29194790	-1.96230
MEXAM	.67406054	1.04126
KIND	.77257832	2.30390
BLKIND	-.72824712	-1.22107
SUMMER	.12395605	.35416
BLKHS	1.02035387	1.74465
MEXHS	-.17377194	-.20681

$R^2 = .1367$

Table A- 20

Effects of Individual Characteristics for Grade 1, Full-Year, Both Parents Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-1.2233	-1.384
CHILD	-.35542	-2.719
INCOME	.40667-03	3.292
AGE	1.4445	2.515
MHSG	-.29106	-.262
MSOHS	-.55237	-.483
M79	-1.2297	-.966
M06	-.51235	-.327
MOCLER	-2.9477	-1.394
MOSKIL	-.30075	-.401
MOSEMI	-1.1387	-.703
MOUNSK	-1.8022	-1.161
FEMALE	.38445	.748
RURAL	1.1677	1.027
FHSG	-1.7300	-1.895
FSOHS	-.94991	-.985
F79	-2.1581	-1.993
F06	-1.0700	-.950
FACLER	1.2541	.522
FASKIL	.57558	.491
FASEMI	-.22140	-.191
FAUNSK	-.40318-01	-.034
KIND	3.4730	3.935
BLACK	.81015	.561
MEXAM	-.46949	-.328
BLCKHS	2.0395	1.785
MEXHS	.80789-01	.054
BLKIND	-3.5649	-3.107
CONSTANT	11.996	2.822
R^2	.3736	
N	206	

Table A- 21

Effects of Individual Characteristics for Grade 1, Full-Year, Both Parents Sample on Child's MRT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HOSTRT	-.25620	-.482
CHILD	-.19927	-2.533
INCOME	.20752-03	2.791
AGE	.73170	2.117
MHSG	-.33519	-.502
MSOHS	-.36801	-1.408
M79	-1.0457	-1.365
M06	-1.7011	-1.801
MOCLER	-2.0419	-1.604
MOSKIL	-2.0940	-1.550
MOSEMI	-1.9914	-2.044
MOUNSK	-2.1058	-2.254
FEMALE	.53791	1.739
RURAL	.83016	1.213
FHSG	-.53410	-1.155
FSOHS	-.32979	-.568
F79	-.22862	-.351
F06	-.33882	-.500
FACLER	2.7265	1.885
FASKIL	1.3179	1.869
FASEMI	.44165	.633
FAUNSK	.21654	.306
KIND	-.30856-01	-.058
BLACK	-2.0378	-2.343
MEXAM	-1.9053	-2.209
BLCKHS	1.1839	1.722
MEXHS	.51933	.573
BLKIND	.29301	.424
CONSTANT	7.3912	2.886
R^2	.4446	
N	206	

Table A-22

Effects of Individual Characteristics for Grade 1, Full-Year, Mother Only Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	3.8599	1.729
CHILD	-.60990	-2.169
INCOME	-.66600-03	-1.327
AGE	4.2986	4.034
MHSG	-4.9458	-2.678
MSOHS	-5.2422	-2.993
M79	-5.8362	-3.208
MO6	-7.1247	-2.771
MOCLER	1.7819	.350
MOSKIL	-.20731	-.039
MOSEMI	.75493	.165
MOUNSK	2.3408	.507
FEMALE	1.1408	1.125
RURAL	3.4721	1.617
KIND	5.2105	1.680
BLACK	4.2629	1.090
BLCKHS	-2.5309	-1.010
SEPAR	-.23355	-.169
WIDOW	1.0035	.622
NEVMAR	-1.1734	-.730
BLKIND	-3.1243	-.937
CONSTANT	-6.3158	-.840

R^2	.5367
N	67

Table A- 23

Effects of Individual Characteristics for Grade 1, Full-Year, Mother Only Sample on Child's MRT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	3.7510	2.694
CHILD	-.81014-01	-.461
INCOME	-.10494-02	-3.341
AGE	3.4948	5.237
MHSG	-3.0977	-2.735
MSOHS	-3.5834	-3.268
M79	-4.2030	-3.151
M36	-5.4793	-3.404
MOCLER	-2.4925	-.772
MOSKIL	-2.7314	-.818
MOSEMI	-5.2880	-1.841
MOUNSK	-3.4288	-1.187
FEMALE	1.2244	1.929
RURAL	2.6556	1.975
KIND	4.6738	2.407
BLACK	3.3953	1.382
BLCKHS	-2.5096	-1.600
SEPAR	.20386	.236
WIDOW	.11148	.110
NEVMAR	.21940	.217
BLKIND	-1.9415	-.930
CONSTANT	-9.2064	-1.957

R^2 .6055

N 67

Table A- 24

Effects of Individual Characteristics for Grade 1, Summer, Both Parents Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-.36056	-1.304
CHILD	-.17439	-1.895
INCOME	-.39910-05	-.173
AGE	.76504	2.791
MHSG	-.19389	-.345
MSOHS	-.50813	-.869
M79	-.73633	-1.145
M06	-1.6222	-1.974
MOCLER	2.4232	1.977
MOSKIL	1.8756	1.509
MOSEMI	.15849	.149
MOUNSK	.50559	.518
FEMALE	-.12013	-.433
RURAL	.24036	.701
FHSG	.34171	.691
FSOHS	-.32696-02	-.006
F79	-.47526	-.890
F06	-.90463	-1.453
FACLER	-.92084	-1.079
FASKIL	-1.0059	-1.633
FASEMI	-1.1680	-1.895
FAUNSK	-1.6354	-2.573
KIND	.96199	2.504
BLACK	-2.1029	-3.112
MEXAM	-1.1215	-1.688
BLCKHS	1.9970	3.125
MEXHS	1.5323	1.777
BLKIND	-.59329	-.895
CONSTANT	16.562	9.420

R^2	.1860
N	623

Table A-25

Effects of Individual Characteristics for Grade 1, Summer, Both Parents Sample on Child's MRT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-.24471	-.937
CHILD	-.73435-01	-1.425
INCOME	.29696-04	.707
AGE	.67353	3.376
MHSG	-.19519	-.477
MSOHS	-.52500	-1.472
M79	-.92153	-1.971
MO6	-1.5461	-2.588
MOCLER	1.7928	2.017
MOSKIL	.86361	.956
MOSEMI	.77455	.998
MOUNSK	.37256	1.231
FEMALE	.41939	2.079
RURAL	.22373	.898
FHSG	.59070	1.643
FSOHS	.28363	.749
F79	.44287	1.141
FO6	-.10078	-.223
FACLER	-.54639	-.891
FASKIL	-.79493	-1.776
FASEMI	-1.0976	-2.450
FAUNSK	-1.2260	-2.654
KIND	.77003	3.078
BLACK	-1.5515	-3.159
MEXAM	-.72446	-1.500
BLCKHS	.95437	2.065
MEXHS	.79610	1.270
BLKIND	-.57811	-1.200
CONSTANT	5.4867	3.815
R^2	.2038	
N	623	

Table A-26

Effects of Individual Characteristics for Grade 1, Summer, Mother Only Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	2.6239	2.393
CHILD	.19624-01	.116
INCOME	.36159-04	.455
AGE	1.6469	2.519
MHSG	1.6421	.653
MSOHS	1.0900	.442
M79	.69879-01	.028
M06	.12619	.043
MOCLER	-.27340	-.081
MOSKIL	-.14436	-.038
MOSEMI	-1.9988	-.632
MOUNSK	-2.5489	-.892
FEMALE	-.50567	-.722
RURAL	-.23003	-.201
KIND	1.2242	1.051
BLACK	2.2136	1.474
BLCKHS	-3.4442	-2.491
SEPAR	-1.0608	-1.242
WIDOW	-.13617	-.108
NEVMAR	-.45422	-.423
BLKIND	-.79632	-.520
CONSTANT	8.7391	1.799
R^2	.1898	
N	143	

Table A-27

Effects of Individual Characteristics for Grade 1, Summer, Mother Only Sample on Child's MRT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	.76545	1.048
CHILD	-.29556-01	-.282
INCOME	.20548-03	1.462
AGE	1.1211	2.572
MHSG	1.0896	.650
MCOHS	1.2362	.758
M79	.76562	.467
MDE	.16575	.086
MOCLER	-3.1094	-1.384
MOSKIL	-3.9920	-1.587
MOSEMI	-4.3797	-2.200
MOUNSK	-4.4257	-2.299
FEMALE	.31486	.675
RURAL	-.54605	-.715
KIND	1.3410	1.726
BLACK	.35924	.359
BLCKHS	-1.0794	-1.171
SEPAR	.27339	.480
WIDOW	.21280	.253
NEVMAR	-.80143	-1.121
BLKIND	-1.1204	-1.098
CONSTANT	3.6114	1.115

R^2	.2246	
N	143	

Table A-28

Effects of Individual Characteristics for Grade 2, Full-Year, Both Parents Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-.52404	-.492
CHILD	-.22459-01	-.162
INCOME	.21502-03	1.864
AGE	-.11931	-.224
MHSG	.62977	.591
MSOHS	-.29219	-.268
M79	.75975	.646
M06	1.2507	.352
MOCLER	-2.7688	-1.043
MOSKIL	-2.1731	-.995
MOSEMI	-3.0774	-1.706
MOUNSK	-3.4185	-2.058
FEMALE	-.25510	-.491
RURAL	-1.3435	-1.810
FHSG	-1.1353	-1.323
FSOHS	-1.4368	-1.509
F79	-1.8712	-1.835
F06	-2.2117	-1.327
FACLER	-.27316	-.143
FASKIL	.55903	.493
FASEMI	.10973	.100
FAUNSK	.50967-01	.046
KIND	3.5650	3.744
BLACK	-2.1368	-1.952
MEXAM	-1.2875	-.874
BLCKHS	.60220	.484
MEXHS	1.7758	1.006
BLKIND	-2.8007	-2.524
CONSTANT	27.194	6.381

R^2	.3367
N	218

Table A-29

Effects of Individual Characteristics for Grade 2, Full-Year, Both Parents Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-2.1671	-.211
CHILD	.38388	.288
INCOME	.45105-03	.406
AGE	5.6230	1.096
MHSG	-11.216	-1.073
MSOHS	-24.172	-2.300
M79	-4.7757	-.421
M06	-2.9571	-.209
MOCLER	-27.960	-1.093
MOSKIL	-25.933	-1.231
MOSEMI	-37.632	-2.164
MOUNSK	-45.080	-2.803
FEMALE	2.1399	.439
RURAL	-5.5290	-.772
FHSG	-13.081	-1.581
FSOHS	-17.544	-1.911
F79	-25.485	-2.592
F06	-24.453	-2.096
FACLER	-13.397	-.727
FASKIL	1.6829	.154
FASEMI	-6.5434	-.625
FAUNSK	-5.5849	-.614
KIND	13.852	1.509
BLACK	-35.536	-3.376
MEXAM	-19.563	-1.377
BLCKHS	10.496	.875
MEXHS	-2.5425	-.149
BLKIND	-10.375	-.970
CONSTANT	215.05	5.234

R^2 .3389

N 218

Table A-30

Effects of Individual Characteristics for Grade 2, Full-Year, Mother Only Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HOSTRT	1.1421	.658
CHILD	.22981	.855
INCOME	-.39213-05	-.032
AGE	.41317-01	.045
MHSG	2.1185	1.052
MSOHS	.99619	.495
M79	.98357	.374
MOE	.18342	.061
MOCLER	-3.8259	-1.059
MOSKIL	-3.9597	-1.217
MOSEMI	-7.3354	-2.503
MOUNSK	-5.2714	-2.009
FEMALE	.38634	.385
RURAL	-.35748	-.173
KIND	1.9971	.630
BLACK	-3.6577	-1.119
BLCKHS	.52122	.259
SEPAR	.53123	.430
WIDOW	3.7390	1.748
NEVMAR	1.1060	.576
BLKIND	-.69418	-.212
CONSTANT	24.824	3.421
R^2	.4283	
N	75	

Table A-31

Effects of Individual Characteristics for Grade 2, Full-Year, Mother Only Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HOSTRT	-10.413	-.576
CHILD	2.3849	.895
INCOME	-.24979-03	-.081
AGE	2.5760	.280
MHSG	11.696	.576
MCOHS	-13.930	-.686
M79	-9.3579	-.394
MC6	-30.085	-.995
MOCLER	-51.440	-1.413
MOSKIL	-2.2672	-.069
MOSEMI	-40.660	-1.377
MOUNSK	-37.061	-1.402
FEMALE	4.3917	.435
RURAL	-5.2846	-.254
KIND	6.7393	.212
BLACK	-35.871	-1.119
BLCKHS	25.097	1.238
SEPAR	-7.5644	-.608
WIDOW	19.977	1.266
NEVMAR	7.4481	.385
BLKIND	-17.218	-.522
CONSTANT	137.50	2.702

R^2	.3753
N	75

Table A-32

Effects of Individual Characteristics for Grade 2, Summer, Both Parents Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-.48079	-1.320
CHILD	-.11495	-1.609
INCOME	.19946-03	3.574
AGE	.18573	.699
MHSG	-.13008	-.188
MSOHS	-.32200	-1.167
M79	-.58091	-.939
M06	-.53014	-.627
MOCLER	.25031	.163
MOSKIL	.62613	.414
MOSEMI	.93134-02	.007
MOUNSK	.37588	.286
FEMALE	-.71379	-2.535
RURAL	-.30014-01	-.246
FHSG	-.21258	-.433
FSOHS	-.30104-01	-.060
F79	-.95179-01	-.176
F05	-.36859	-1.452
FACLER	.24387	.277
FASKIL	-.66888	-.990
FASEMI	-1.0394	-1.547
FAUNSK	-1.4097	-2.052
KIND	1.4615	4.075
BLACK	-.48765	-.826
MEXAM	-1.1453	-1.763
BLCKHS	-.86583-01	-.130
MEXHS	-.66409	-.771
BLKIND	-1.5613	-2.381
CONSTANT	22.729	9.727

R^2	.2057
N	635

Table A-33

Effects of Individual Characteristics for Grade 2, Summer, Both Parents Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HOSTRT	-11.535	-2.629
CHILD	.22578	.262
INCOME	.14627-02	2.176
AGE	-.54480	-.201
MHSG	-5.0985	-.612
MSOHS	-23.627	-2.784
M79	-17.990	-2.059
MO6	-17.932	-1.767
MOCLER	5.6490	.305
MOSKIL	-3.8777	-.213
MOSEMI	.22917	.014
MOUNSK	-.47354	-.030
FEMALE	12.647	3.702
RURAL	2.1626	.551
FHSG	8.6016	1.455
FSOHS	7.3220	1.207
F79	-2.5416	-.391
FO6	1.3231	.184
FACLER	-3.5098	-.330
FASKIL	-13.113	-1.611
FASEMI	-11.265	-1.392
FAUNSK	-17.203	-2.079
KIND	9.4449	1.955
BLACK	-14.939	-2.086
MEXAM	-9.2329	-1.180
BLCKHS	6.5143	.809
MEXHS	1.4635	.141
BLKIND	-10.070	-1.275
CONSTANT	193.35	6.869

R^2	.1857
N	635

Table A-34

Effects of Individual Characteristics for Grade 2, Summer, Mother Only Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	1.1149	1.371
CHILD	-.54185	-3.605
INCOME	.59878-03	3.255
AGE	.27980	.436
MHSG	.57319	.409
MSOHS	-.55388	-.396
M79	-1.0831	-.748
MO6	-1.6920	-.918
MOCLER	6.7934	2.260
MOSKIL	6.7633	2.399
MOSEMI	6.2652	2.625
MOUNSK	6.0240	2.555
FEMALE	-1.1261	-1.911
RURAL	.41082	.467
KIND	1.9402	2.105
BLACK	-.84905	-.696
BLCKHS	-.60232	-.538
SEPAR	-1.0101	-1.494
WIDOW	.32135	.849
NEVMAR	-.95905	-.861
BLKIND	-.33992	-.283
CONSTANT	14.788	2.522

R^2	.4125
N	134

Table A-35

Effects of Individual Characteristics for Grade 2, Summer, Mother Only Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	11.379	1.000
CHILD	-4.3799	-2.082
INCOME	.13568-02	.536
AGE	8.7208	.972
MHSG	9.0377	.461
MSOHS	6.1080	.312
M79	-9.0464	-.397
MC6	-1.8308	-.071
MOCLER	42.019	1.000
MOSKIL	42.779	1.084
MOSEMI	45.771	1.400
MOUNSK	42.140	1.489
FEMALE	1.0276	.125
RURAL	15.424	1.252
KIND	15.144	1.174
BLACK	-13.583	-1.090
BLCKHS	.40976	.026
SEPAR	12.186	1.288
WIDOW	4.0271	.297
NEVMAR	-3.7998	-.271
BLKIND	-5.8958	-.350
CONSTANT	56.650	.690
R^2	.2236	
N	134	

Table A-36

Effects of Individual Characteristics for Grade 3, Summer, Both Parents Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	-.21363-01	-.045
CHILD	-.25761	-2.684
INCOME	.14209-03	1.749
AGE	.16355	.483
MHSG	-.94039	-.895
MSOHS	-1.8057	-1.609
M79	-2.8776	-2.482
MO6	-3.7334	-2.791
MOCLER	-2.3055	-1.292
MOSKIL	-.30701	-.146
MOSEMI	-1.2781	-.821
MOUNSK	-.94501	-.572
FEMALE	.66151-01	.168
RURAL	-.39608	-.791
FHSG	-.92746	-1.219
FSOHS	-1.3200	-1.669
F79	-.37368	-.460
FO6	-.70406	-.814
FACLER	-.77145	-.626
FASKIL	.10041	.101
FASEMI	-.30312	-.318
FAUNSK	-.37564	-.384
KIND	1.9739	3.455
BLACK	-1.7737	-2.270
MEXAM	-1.7172	-1.459
BLCKHS	.60324	.694
MEXHS	-.15018	-.097
BLKIND	-.25920	-.302
CONSTANT	29.366	9.345
R^2	.2673	
N	426	

Table A-37

Effects of Individual Characteristics for Grade 3, Summer, Both Parents Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HOSTRT	-.54596	-.077
CHILD	-.91491	-.653
INCOME	.24194-02	2.040
AGE	-4.5748	-.898
MHSG	-23.017	-1.495
MSOHS	-39.535	-2.358
M79	-54.002	-3.191
M06	-52.384	-2.684
MOCLER	-4.3793	-.168
MOSKIL	9.7725	.318
MOSEMI	-20.024	-.881
MOUNSK	-12.899	-.597
FEMALE	19.216	3.173
RURAL	3.8030	.520
FHSG	-26.359	-2.375
FSOHS	-29.475	-2.554
F79	-24.318	-2.051
F06	-29.273	-2.240
FACLER	11.389	.633
FASKIL	2.1584	.149
FASEMI	-2.5192	-.178
FAUNSK	-1.5815	-.111
KIND	9.6962	1.163
BLACK	-19.393	-1.613
MEXAM	-45.391	-2.698
BLCKHS	-13.861	-1.092
MEXHS	13.369	.857
BLKIND	-11.706	-.901
CONSTANT	341.42	6.882
R^2	.2895	
N	426	

Table A-38

Effects of Individual Characteristics for Grade 3, Summer, Mother Only Sample on Child's ITPA Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	1.5192	1.137
CHILD	-.34123-01	-.504
INCOME	.27087-03	.995
AGE	.32697	.523
MHSG	1.3692	.705
MSOHS	1.1111	.574
M79	.53230	.266
MC5	1.0003	.439
MOCLER	-5.3330	-1.827
MOSKIL	-3.4432	-.883
MOSEMI	-3.9254	-1.505
MOUNSK	-4.0383	-1.585
FEMALE	.14022	.182
RURAL	-.31024	-.279
KIND	1.2323	.921
BLACK	-.38160	-.636
BLCKHS	-2.0084	-1.265
SEPAR	.71542	.632
WIDOW	.41001-01	.033
NEVMAR	.99974	.679
BLKIND	-.70691	-.432
CONSTANT	24.380	4.095

R^2	.1873
N	114

Table A-39

Effects of Individual Characteristics for Grade 3, Summer, Mother Only Sample on Child's SAT Score

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
HDSTRT	8.3131	.547
CHILD	.60917	.287
INCOME	.34655-02	1.119
AGE	1.8914	.266
MHSG	9.3541	.424
MSOHS	1.9923	.090
M79	.35702	.042
MO6	27.260	1.050
MOCLER	-53.267	-1.904
MOSKIL	-102.34	-2.306
MOSEMI	-69.525	-2.342
MOUNSK	-63.774	-2.200
FEMALE	19.143	1.837
RURAL	23.610	1.863
KIND	1.7294	.114
BLACK	-15.911	-.905
BLCKHS	-12.592	-.697
SEPAR	14.778	1.147
WIDOW	2.9034	.208
NEVMAR	11.772	.709
BLKIND	-2.0156	-.109
CONSTANT	237.33	3.503

R^2	.2753	
N	114	

Table A-40

Discriminant Regression Results for Grade 1, Full-Year, Both Parents Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.24566-07	-.130
INCOME	-.25912-04	-1.452
AGE	-.19436-01	-.234
MHSG	.19475	1.214
MSOHS	.13117	.800
M79	.31616-01	.173
MO6	-.71141-01	-.316
MOCLER	.17706-01	.058
MOSKIL	.13406	.412
MOSEMI	.30117-01	.382
MOUNSK	.99206-01	.439
FEMALE	-.21017-01	-.281
RURAL	.87519-01	.579
FHSG	.28332	2.215
FSOHS	.35502	2.573
F79	.33403	2.143
FO6	.33422	2.074
FACLER	-.64308	-1.888
FASKIL	-.22581	-1.336
FASEMI	-.24130	-1.435
FAUNSK	-.13375	-.783
KIND	.27038-01	.324
BLACK	.16619-01	.124
MEXAM	.75893-01	.484
CONSTANT	.43294	.730
R^2	.1004	
F-statistic (24, 181)	.8417	
N	206	

Table A-41

Discriminant Regression Results for Grade 1, Full-Year, Mother Only Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.71987-01	2.015
INCOME	.49475-06	.007
AGE	-.22277-01	-.154
MHSG	-.74184-02	-.031
MSOHS	-.16458	-.714
M79	-.20711	-.765
MC6	.27684	.798
MOCLER	-.35020	-1.488
MOSKIL	-.76200	-1.126
MOSEMI	.14824	.250
MOUNSK	-.87637-01	-.150
FEMALE	.23229	1.734
RURAL	.15240	.522
KIND	-.87605-01	-.493
BLACK	-.32719-01	-.159
SEPAR	-.35339	-1.954
WIDOW	-.39312	-1.867
NEVMAR	-.39481	-1.905
CONSTANT	.83480	.858

R^2	.4133
F-statistic (18, 48)	1.879
N	67

Table A-42

Discriminant Regression Results for Grade 1, Summer, Both Parents Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.34175-01	3.382
INCOME	-.17586-04	-2.117
AGE	-.55185-01	-1.420
MHSG	-.11562	-1.428
MSOHS	.52477-01	.622
M79	-.91560-01	-.987
MO6	-.41301-01	-.349
MOCLER	-.27682	-1.578
MOSKIL	-.75289-01	-.421
MOSEMI	-.45314-01	-.299
MOUNSK	-.21999	-1.574
FEMALE	-.14491-02	-.036
RURAL	.50330-01	1.222
FHSG	-.57395-01	-.814
FSOHS	-.93608-01	-1.115
F79	.49333-01	.641
FO6	-.40926-02	-.046
FACLER	.17113	1.392
FASKIL	.11225	1.268
FASEMI	.13541	1.527
FAUNSK	.22010	2.416
KIND	.57606-01	1.319
BLACK	-.21219-01	-.414
MEXAM	.26578-01	.391
CONSTANT	.89432	3.165

R^2	.1017	
F-statistic (24, 598)	2.821	
N	623	

Table A-43

Discriminant Regression Results for Grade 1, Summer, Mother Only Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.77427-02	-.340
INCOME	-.19469-04	-.682
AGE	-.17239-01	-.199
MHSG	-.21203-01	-.062
MSOHS	-.54649-01	-.196
M79	.51639-01	.155
M06	.11182	.285
MOCLER	-.17884	-.393
MOSKIL	-.50156	-1.194
MOSEMI	-.22363	-.555
MOUNSK	-.42909	-1.103
FEMALE	-.33904-01	-.920
RURAL	.90953-01	.587
KIND	.12765	1.208
BLACK	-.11439	-1.113
SEPAR	.46920-01	.408
WIDOW	.12939	.763
NEVMAR	.18516	1.285
CONSTANT	1.1101	1.779

R^2	.0951
F-statistic (18, 124)	.7240
N	143

Table A-44

Discriminant Regression Results for Grade 2, Full-Year, Both Parents Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.41985-01	2.230
INCOME	-.11890-04	-.747
AGE	-.13865	-1.924
MHSG	-.12339	-.832
MSOHS	-.46454-01	-.310
M79	-.17749	-1.107
M06	-.41302	-2.072
MOCLER	.22156	.610
MOSKIL	-.28127	-.934
MOSEMI	-.39281-02	-.016
MOUNSK	-.64707-01	-.281
FEMALE	.52855-01	.880
RURAL	-.30430-01	-.797
FHSG	-.82907-01	-.700
FSOHS	.75964-01	.592
F79	-.37028-01	-.619
F06	-.13364	-.803
FACLER	.35474-01	.326
FASKIL	.57655-01	.369
FASEMI	.10136	.678
FAUNSK	.10408	.681
KIND	-.44519-01	-.522
BLACK	.12515-01	.124
MEXAM	.25737	1.799
CONSTANT	1.4583	2.525

R^2	.1274
F-statistic (24, 193)	1.174
N	218

Table A-45

Discriminant Regression Results for Grade 2, Full-Year, Mother Only Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.77916-01	-2.466
INCOME	.26219-04	.692
AGE	-.28019	-2.577
MHSG	-.10457-01	-.041
MSOHS	.16699	.667
M79	-.43002-01	-.147
MO6	.37454	1.030
MOCLER	-.15330	-.347
MOSKIL	.32520	.811
MOSEMI	.40766	1.161
MOUNSK	.24582	.779
FEMALE	-.40433-01	-.320
RURAL	.98359-01	.378
KIND	-.12269-01	-.070
BLACK	.12120	.822
SEPAR	-.10197	-.678
WIDOW	.20771	1.059
NEVMAR	-.13074	-.551
CONSTANT	2.4704	3.151

R^2	.3413
F-statistic (18, 56)	1.612
N	75

Table A-46

Discriminant Regression Results for Grade 2, Summer, Both Parents Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.17470-01	1.737
INCOME	-.58438-05	-.740
AGE	-.11769	-3.155
MHSG	.18582	1.910
MSOHS	.70395-01	.707
M79	.15082	1.471
M06	.32903-01	.777
MOCLER	-.44071	-2.033
MOSKIL	-.32210	-1.512
MOSEMI	-.41733	-2.151
MOUNSK	-.32222	-1.736
FEMALE	-.52748-01	-1.316
RURAL	-.26934-02	-.058
FHSG	-.26999-01	-.388
FSOHS	.35883-01	.506
F79	.17343-01	.234
F06	.45350-01	.535
FACLER	.10240	.821
FASKIL	.32236-01	.337
FASEMI	.61939-01	.651
FAUNSK	.68581-01	.705
KIND	-.20489-02	-.047
BLACK	-.68795-02	-.141
MEXAM	.24189-01	.369
CONSTANT	1.4290	4.383

R^2	.0533
F-statistic (24, 610)	1.432
N	635

Table A-47

Discriminant Regression Results for Grade 2, Summer, Mother Only Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.17497-01	.687
INCOME	.19175-04	.627
AGE	-.11790	-1.108
MHSG	.58938-01	.287
MSOHS	.79480-01	.327
M79	-.24399-01	-.341
MO6	.14097	.454
MOCLER	.39873	.781
MOSKIL	.17104	.354
MOSEMI	.35707	.870
MOUNSK	.28945	.722
FEMALE	.40056-01	.400
RURAL	.39914-02	.060
KIND	-.10958	-1.046
BLACK	.53485-02	.047
SEPAR	.16314	1.467
WIDOW	.14683	.913
NEVMAR	.17162	1.012
CONSTANT	.97693	.894
R^2	.0768	
F-statistic (18, 115)	.5314	
N	134	

Table A-48

Discriminant Regression Results for Grade 3, Summer, Both Parents Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.55542-02	.450
INCOME	.34157-05	.326
AGE	-.17529	-4.004
MHSG	.14453	1.057
MSOHS	.15655	1.089
M79	.80392-01	.545
MO6	.16633	.979
MOCLER	.12537	.546
MOSKIL	-.36834	-1.369
MOSEMI	-.98554-01	-.495
MOUNSK	-.15371	-.834
FEMALE	.21496-02	.043
RURAL	.11765-01	.183
FHSG	.42231-01	.433
FSOHS	.11031	1.086
F79	.57080-01	.548
FO6	.17973	1.621
FACLER	-.32673-01	-.520
FASKIL	-.37392-01	-.297
FASEMI	-.10176	-.818
FAUNSK	-.82428-01	-.657
KIND	-.54894-01	-.898
BLACK	.96096-01	1.578
MEXAM	-.64059-02	-.061
CONSTANT	1.8770	4.433

R^2	.0754
F-statistic (24, 401)	1.362
N	426

Table A-49

Discriminant Regression Results for Grade 3, Summer, Mother Only Sample, with Head Start as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.56567-02	.268
INCOME	-.28447-04	-.918
AGE	-.10662	-1.272
MHSG	.77528-01	.299
MSOHS	.57268-01	.220
M79	.38735-01	.144
MO6	-.29703	-.974
MOCLER	.70484	1.829
MOSKIL	1.0060	1.968
MOSEMI	.53499	1.566
MOUNSK	.54621	1.639
FEMALE	-.78273-01	-.754
RURAL	-.17874	-1.239
KIND	-.24291	-2.175
BLACK	.78619-01	.673
SEPAR	-.19037	-1.293
WIDOW	.40183-01	.245
NEVMAR	-.20489	-1.043
CONSTANT	1.0752	1.368
R^2	.1802	
F-statistic (18, 95)	1.160	
N	114	

Table A-50

Discriminant Regression Results for Grade 1, Full-Year, Both Parents Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.73420-02	-.439
INCOME	.31184-04	1.987
AGE	-.55207-01	-.750
MHSG	-.33172	-2.361
MSOHS	-.18526	-1.277
M79	-.34395	-2.146
M06	-.54076	-2.759
MOCLER	.32023	1.187
MOSKIL	.19429-01	.067
MOSEMI	.25165	1.203
MOUNSK	.20168	1.007
FEMALE	.30603-01	.461
RURAL	-.59953	-4.702
FHSG	-.51170-01	-.441
FSOHS	-.52891-01	-.431
F79	-.85309-01	-.616
F06	-.19061	-1.337
FACLER	.25162	.825
FASKIL	.47535-01	.316
FASEMI	.33439-01	.224
FAUNSK	.24210-01	.159
BLACK	-.32600	-2.797
MEXAM	-.26349	-1.883
CONSTANT	1.2177	2.343

R^2	.2694	
F-statistic (22, 182)	2.918	
N	206	

Table A-51

Discriminant Regression Results for Grade 1, Full-Year, Mother Only Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.11343-01	-.395
INCOME	.20750-03	4.557
AGE	-.15663	-1.371
MHSG	.35863	1.906
MSOHS	.30973	1.720
M79	.27169	1.269
M06	.26168	.946
MOCLER	1.2824	2.673
MOSKIL	1.4685	2.925
MOSEMI	1.1557	2.583
MOUNSK	1.0804	2.434
FEMALE	-.25911	-2.674
RURAL	-.15782	-.654
BLACK	-.19917	-1.224
SEPAR	.37185-01	.601
WIDOW	-.11911	-.757
NEVMAR	-.76947-01	-.462
CONSTANT	.77902-01	.099

R^2	.4700
F-statistic (17, 49)	2.556
N	67

Table A-52

Discriminant Regression Results for Grade 1, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.25580-01	2.713
INCOME	.39447-04	5.193
AGE	.45723-01	1.239
MHSG	.29545-01	.391
MSOHS	-.34574-02	-.120
M79	-.35643-01	-1.104
M06	-.74083-01	-.869
MOCLER	.23571	1.440
MOSKIL	.45137	2.720
MOSEMI	.27875	1.956
MOUNSK	.25513	1.967
FEMALE	-.92996-02	-.250
RURAL	-.15241	-2.333
FHSG	-.46019-01	-.891
FSOHS	-.10913	-1.545
F79	-.14241	-1.998
F06	-.19642	-2.348
FACLER	.15293	1.333
FASKIL	.95740-02	.116
FASEMI	.17205-01	.208
FAUNSK	-.33920-01	-.397
BLACK	.12986	2.725
MEXAM	.30167	4.835
CONSTANT	-.15129	-.573

R^2	.1663
F-statistic (23, 599)	5.1967
N	623

Table A-53

Discriminant Regression Results for Grade 1, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.26647-01	-1.393
INCOME	.39082-04	1.636
AGE	-.12031	-1.658
MHSG	.56666	1.997
MSOHS	.40164	1.443
M79	.34737	1.240
M06	.25734	.776
MOCLER	-.37660	-.873
MOSKIL	-.53337	-1.259
MOSEMI	-.45549	-1.346
MOUNSK	-.23605	-.718
FEMALE	-.25211-01	-.327
RURAL	-.25132	-1.947
BLACK	.13955	1.622
SEPAR	-.50227-01	-.517
WIDOW	-.14430	-1.009
NEVMAR	-.20052	-1.663
CONSTANT	1.3127	2.550

R^2	.2007
F-statistic (17, 125)	1.847
N	143

Table A-54

Discriminant Regression Results for Grade 2, Full-Year, Both Parents Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.20266-01	1.284
INCOME	.54920-04	4.283
AGE	-.11749	-1.956
MHSG	.54318-01	.435
MSOHS	.15304	1.219
M79	-.17081-01	-.096
MC6	.16377-01	.098
MOCLER	.16753	.548
MOSKIL	-.14427	-.570
MOSEMI	-.15360-01	-.074
MOUNSK	.78229-01	.404
FEMALE	.13021	2.192
RURAL	-.43721	-5.537
FHSG	-.41539-01	-.417
FSOHS	-.94777-01	-.864
F79	.51045-01	.432
FC6	-.86971-01	-.621
FACLER	-.17955	-.815
FASKIL	.21617	1.654
FASEMI	.35495-01	.688
FAUNSK	.77700-01	.604
BLACK	-.23543	-2.816
MEXAM	-.43182	-3.712
CONSTANT	1.0020	2.084

R^2	.3753
F-statistic (23, 194)	5.067
N	218

Table A-55

Discriminant Regression Results for Grade 2, Full-Year, Mother Only Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.12980-01	.539
INCOME	.77497-04	2.891
AGE	.74479-01	.908
MHSG	-.29540	-1.576
MSOHS	-.26358	-1.409
M79	-.54092	-2.576
MOE	-.47136	-1.595
MOCLER	.30470	.914
MOSKIL	-.24364	-.824
MOSEMI	.25376	.959
MOUNSK	.15978	.669
FEMALE	.14089	1.496
RURAL	-.53003	-4.024
BLACK	-.21799-01	-.195
SEPAR	-.22223	-2.014
WIDOW	-.17785	-1.209
NEVMAR	-.35444	-2.039
CONSTANT	.21128	.355

R^2	.5782
F-statistic (17, 57)	4.597
N	75

Table A-56

Discriminant Regression Results for Grade 2, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.24756-01	2.695
INCOME	.30927-04	4.329
AGE	.45701-01	1.324
MHSG	.16763	1.882
MSOHS	.70335-01	.769
M79	.46591-01	.495
M06	.44699-01	.407
MOCLER	-.35065	-1.766
MOSKIL	-.37335	-1.914
MOSEMI	-.60121-01	-.337
MOUNSK	-.14981	-.873
FEMALE	.14337-01	.407
RURAL	-.18720	-4.497
FHSG	-.27638-01	-.433
FSOHS	-.21390-01	-.337
F79	-.11391	-1.701
F06	-.19910	-2.571
FACLER	.15150	1.325
FASKIL	-.52964-01	-.716
FASEMI	-.60695-02	-.069
FAUNSK	-.10713	-1.200
BLACK	.77596-01	1.739
MEXAM	.27963	4.733
CONSTANT	.37941-01	.327

R^2	.1982
F-statistic (23, 611)	6.566
N	635

Table A-57

Discriminant Regression Results for Grade 2, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	.43174-01	1.946
INCOME	.25730-04	.953
AGE	.13777-01	.146
MHSG	.26463-01	.124
MSOHS	.22751	1.075
M79	-.14805	-.663
MO8	-.26856	-.980
MOCLER	.67187	1.500
MOSKIL	.46559	1.092
MOSEMI	.41012	1.148
MOUNSK	.37610	1.064
FEMALE	.10141	1.149
RURAL	-.55162-01	-.489
BLACK	.33603-01	.834
SEPAR	-.51656-01	-.608
WIDOW	-.33633	-2.420
NEVMAR	-.17503	-1.172
CONSTANT	-.17708	-.204

R^2	.2603
F-statistic (17, 116)	2.402
N	134

Table A-58

Discriminant Regression Results for Grade 3, Summer, Both Parents Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.32281-02	-.321
INCOME	.29077-04	3.450
AGE	-.53103-02	-.177
MHSG	.57365-02	.051
MSOHS	-.76251-01	-.650
M79	-.59320-01	-.490
MC6	-.16572	-1.198
MCOLER	.20587	1.101
MOSKIL	.15043	.586
MOSEMI	-.25393-01	-.156
MOUNSK	.47724-01	.309
FEMALE	.12245	3.026
RURAL	-.39012	-8.005
FHSG	.62417-01	.785
FSOHS	.90531-01	1.094
F79	-.54766-01	-.763
FC6	-.11785	-1.305
FACLER	.14984	1.156
FASKIL	.27870	2.698
FASEMI	.15367	1.518
FAUNSK	.16591	1.627
BLACK	.17306	3.684
MEXAM	-.26375	-3.091
CONSTANT	.14694	.425

R^2	.3514
F-statistic (23, 402)	9.469
N	426

Table A-59

Discriminant Regression Results for Grade 3, Summer, Mother Only Sample, with Kindergarten as the Dependent Variable

INDEPENDENT VARIABLE	COEFFICIENT	T-RATIO
CHILD	-.36613-01	-1.635
INCOME	.98587-04	2.907
AGE	.94923-01	1.249
MHSG	.31706	1.349
MSOHS	.26163	1.105
M79	.17920	.728
MO6	.25412-01	.091
MOCLER	.38481	1.100
MOSKIL	1.0745	2.366
MOSEMI	.38146	1.232
MOUNSK	.49977	1.661
FEMALE	-.71799-01	-.759
RURAL	-.23409	-1.305
BLACK	.33785-02	.032
SEPAR	-.17612	-1.320
WIDOW	-.35371	-2.429
NEVMAR	-.30252	-1.711
CONSTANT	-.81756	-1.146

R^2	.3048
F-statistic (17, 96)	2.476
N	114

BIBLIOGRAPHY

Althauser, Robert P., and Rubin, Donald. "Measurement Error and Regression to the Mean in Matched Samples." Social Forces 50, No. 2 (December 1971): 206-214.

Barnow, Burt S. "Conditions for the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation." Madison: Institute for Research on Poverty, 1972. Discussion Paper 122-72.

Bereiter, Carl. "Some Persisting Dilemmas in the Measurement of Change." In Problems in Measuring Change, edited by C.W. Harris, pp. 3-20. Madison: The University of Wisconsin Press, 1967.

Bissell, Joan S. Implementation of Planned Variation in Head Start. Washington, D.C.: Office of Child Development, U.S. Department of Health, Education, and Welfare, April 1971.

Blalock, Hubert M. Jr. Causal Inferences in Nonexperimental Research. Chapel Hill: The University of North Carolina Press, 1961.

Bohrnstedt, George W. "Observations on the Measurement of Change." In Sociological Methodology 1969, edited by E.F. Borgatta, pp. 113-133. San Francisco: Jossey-Bass, Inc., 1969.

Bowles, Samuel. "Towards an Educational Production Function." In Education, Income, and Human Capital, edited by W. Lee Hansen. New York: National Bureau of Economic Research, 1970.

Bowles, Samuel S., and Levin, Henry M. "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Evidence." Journal of Human Resources 3, No. 1 (Winter 1968): 3-29.

Brownlee, K.A. Statistical Theory and Methodology in Science and Engineering. 2d ed. New York: John Wiley & Sons, Inc., 1965.

Cain, Glen G., and Hollister, Robinson G. "The Methodology of Evaluating Social Action Programs." Madison: Institute for Research on Poverty, 1969. Discussion Paper 42-69.

Cain, Glen G., and Watts, Harold W. "Problems in Making Inferences from the Coleman Report." American Sociological Review 35, No. 2 (April 1970): 228-242.

Campbell, Donald T. "From Description to Experimentation: Interpreting Trends as Quasi-Experiments." In Problems in Measuring Change, edited by C.W. Harris, pp. 212-244. Madison: The University of Wisconsin Press, 1967.

Campbell, Donald T. "Reforms as Experiments." American Psychologist 24, No. 4 (April 1969): 409-429.

Campbell, Donald T., and Erlebacher, Albert. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 185-210. New York: Brunner/Mazel, 1970.

Campbell, Donald T., and Erlebacher, Albert. "Reply to the Replies." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 221-225. New York: Brunner/Mazel, 1970.

Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Co., 1963.

Cicirelli, Victor G. "Head Start: Brief of the Study." In Britannica Review of American Education. Vol. I., edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Cicirelli, Victor G. "The Relevance of the Regression Artifact Problem to the Westinghouse-Ohio Evaluation of Head Start: A Reply to Campbell and Erlebacher." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 211-215. New York: Brunner/Mazel, 1970.

Cicirelli, Victor G.; Evans, John W.; and Schiller, Jeffry S. "The Impact of Head Start: A Reply to the Report Analysis." Harvard Educational Review 40, No. 1 (Winter 1970): 105-129.

Cicirelli, Victor G. et al. The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. Vol. I and Vol. II. A report presented to the Office of Economic Opportunity pursuant to contract B89-4536, June 1969. Westinghouse Learning Corporation, Ohio University.

Cicirelli, Victor G. et al. "Performances of Disadvantaged Primary-Grade Children on the Revised Illinois Test of Psycholinguistic Abilities." Psychology in the Schools 8, No. 3 (1971): 240-246.

Cochran, W.G. "Errors of Measurement in Statistics." Technometrics 10, No. 4 (November 1968): 637-666.

Coleman, James S. "Reply to Cain and Watts." American Sociological Review 35, No. 2 (April 1970): 242-249.

Coleman, James S. et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Cramer, J.S. "Efficient Grouping, Regression and Correlation in Engel Curve Analysis." Journal of the American Statistical Association 59, No. 305 (March 1964): 233-250.

Cronbach, Lee J., and Furby, Lita. "How We Should Measure 'Change'--Or Should We?" Psychological Bulletin 74, No. 1 (1970): 68-80.

Datta, Lois-Ellin. A Report on Evaluation Studies of Project Head Start. Washington, D.C.: Project Head Start, Office of Child Development, U.S. Department of Health, Education, and Welfare, 1969.

Evans, John W. "Head Start: Comments on the Criticisms." In Britannica Review of American Education. Vol. I., edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Evans, John W., and Schiller, Jeffry. "How Preoccupation with Possible Regression Artifacts Can Lead to a Faulty Strategy for the Evaluation of Social Action Programs: A Reply to Campbell and Erlebacher." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 216-220. New York: Brunner/Mazel, 1970.

Gaito, John, and Wiley, David E. "Univariate Analysis of Variance Procedures in the Measurement of Change." In Problems in Measuring Change, edited by C.W. Harris, pp. 60-84. Madison: The University of Wisconsin Press, 1967.

Goldberger, Arthur S. "Econometrics and Psychometrics: A Survey of Communalities." Psychometrika 36, No. 2 (June 1971): 83-107.

Goldberger, Arthur S. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations." Madison: Institute for Research on Poverty, 1972. Discussion Paper 123-72. (1972a)

Goldberger, Arthur S. "Selection Bias in Evaluating Treatment Effects: The Case of Interaction." Madison: Institute for Research on Poverty, 1972. Discussion Paper 129-72. (1972b)

Grotberg, Edith S. "Review of Research 1965-1969." Washington, D.C.: Office of Economic Opportunity, 1969.

Hanushek, Erik A. "The Education of Negroes and Whites." Ph.D. dissertation, Massachusetts Institute of Technology, 1968.

Herzog, E. et al. "But Some Are Poorer Than Others: SES Differences in a Preschool Program." American Journal of Orthopsychiatry 42, No. 1 (January 1972): 4-22.

Hilderth, Gertrude H.; Griffiths, Nellie L.; and McGauvran, Mary E. *Manual of Directions: Metropolitan Readiness Tests*. U.S.: Harcourt, Brace & World, Inc., 1966.

Hollingshead, August B. *Social Class and Mental Illness*. New York: John Wiley & Sons, Inc., 1958.

Jensen, Arthur R. "The Culturally Disadvantaged and the Heredity-Environment Uncertainty." In *Head Start and Early Intervention*, Vol. 2 of *Disadvantaged Child*, edited by J. Hellmuth, pp. 27-76. New York: Brunner/ Mazel, 1968.

Jensen, Arthur R. "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review* 39, No. 1 (Winter 1969): 1-123.

Jensen, Arthur R. "Can We and Should We Study Race Differences?" In *Compensatory Education: A National Debate*, Vol. 3 of *Disadvantaged Child*, edited by J. Hellmuth, pp. 124-157. New York: Brunner/Mazel, 1970.

Johnston, J. *Econometric Methods*. New York: McGraw-Hill Book Co., 1963.

Kmenta, Jan. *Elements of Econometrics*. New York: The Macmillan Company, 1971.

Lauman, Lydia Fischer. "Effects of Project Head Start, Summer 1965: A Second Look at the Equality of Economic Opportunity Study." Madison: Institute for Research on Poverty, 1969. Discussion Paper 47-69.

Light, Richard J., and Smith, Paul V. "Choosing a Future: Strategies for Designing and Evaluating New Programs." *Harvard Educational Review* 40, No. 1 (Winter 1970): 1-28.

Linn, Robert L.; Werts, Charles E.; and Tucker, Ludyard R. "The Interpretation of Regression Coefficients in a School Effects Model." *Educational and Psychological Measurement* 31, No. 1 (Spring 1971): 85-94.

Levitan, Sar A. *The Great Society's Poor Laws: A New Approach to Poverty*. Baltimore: The Johns Hopkins University Press, 1969.

Lord, Fredrick M. "Large-Sample Covariance Analysis When the Control Variable is Fallible." *Journal of the American Statistical Association* 55 (1960): 307-321.

Lord, Fredrick M. "Elementary Models for Measuring Change." In *Problems in Measuring Change*, edited by C.W. Harris, pp. 21-38. Madison: The University of Wisconsin Press, 1967.

Lord, Fredrick M. "A Paradox in the Interpretation of Group Comparisons." *Psychological Bulletin* 68 (1967): 304-305.

Lord, Fredrick M., and Novick, Melvin R. Statistical Theory of Mental Tests. Reading, MA.: Addison-Wesley, 1968.

Madow, William G. "Project Head Start, a National Evaluation: A Methodological Critique." In Britannica Review of American Education Vol. 1, edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Moore, Raymond S., and Moore, Dennis R. "The Dangers of Early Schooling." Harper's 245, No. 1466 (July 1972): 58-62.

Murphy, Lois Barclay. "Statement to the House Committee on Education and Labor on the Subject of Head Start." Unpublished manuscript, April 1969.

Scheffe, Henry. The Analysis of Variance. New York: John Wiley & Sons, Inc., 1959.

Smith, Marshall S., and Bissell, Joan S. "Report Analysis: The Impact of Head Start." Harvard Educational Review 40, No. 1 (Winter 1970): 51-104.

Stanley, Julian C., ed. Preschool Programs for the Disadvantaged. Baltimore: The Johns Hopkins University Press, 1972.

Stearns, Marian S. Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families. Washington, D.C.: Office of Child Development, U.S. Department of Health, Education, and Welfare, 1971.

Watts, Harold W., and Horner, David L. "The Educational Benefits of Head Start: A Quantitative Analysis." Madison: Institute for Research on Poverty, 1968. Discussion Paper 14-68.

Webster, Harold, and Bereiter, Carl. "The Reliability of Changes Measured by Mental Test Scores." In Problems in Measuring Change, edited by C.W. Harris, pp. 39-59. Madison: The University of Wisconsin Press, 1967.

Weisbrod, Burton A. "Income Redistribution Effects and Benefit-Cost Analysis." In Problems in Public Expenditure Analysis, edited by Samuel B. Chase, Jr. Washington, D.C.: The Brookings Institution, 1968.

Werts, Charles E., and Linn, Robert L. "Analysing ANCOVA with a Fallible Covariate." Educational and Psychological Measurement 31, No. 1 (Spring 1971): 95-104.

White, Sheldon H. "The National Impact Study of Head Start." In Compensatory Education: A National Debate Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 163-184. New York: Brunner/Mazel, 1970.

Wholey, Joseph S. et al. Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, D.C.: The Urban Institute, 1971.

Williams, Walter, and Evans, John W. "The Politics of Evaluation: The Case of Head Start." The Annals of the American Academy 385 (September 1969): 118-132.